Are countries reforming for the better? A comparison of international bank regulation

Master Thesis
in partial fulfillment of the requirements for the degree of
Master of Science in Volkswirtschaftslehre

at the School of Business and Economics,
Humboldt University of Berlin

presented by
Jan Henner

Berlin, 11.06.2014

Supervisor:
Dr. Sigbert Klinke
(Ladislaus von Bortkiewicz Chair of Statistics)

Examiners:
Prof. Dr. Wolfgang Härdle
(Ladislaus von Bortkiewicz Chair of Statistics)
Prof. Dr. Kerstin Bernoth
(DIW Berlin)

Contents

List of Figures

List of Tables

Abbreviations

GMM – generalized method of moments
IVs – instrumental variables
MIC – mean interitem correlation
NPL – Bank nonperforming loans
PCSE – panel-corrected standard errors
ROA – return on assets

Abstract

I investigate the effect of the Basel accords' pillars on bank stability. The aggregation of the answers in the World Bank dataset on bank regulation provided by Barth et al. (2013b) is shown to be valid only under strict assumptions. A cross-section model provides evidence that a more powerful supervision decreases stability in corrupt countries. The PCSE approach in Boudriga et al. (2009) is shown to be debatable. I apply GMM estimators to tackle persistence and endogeneity modeling nonperforming loans.
Keywords: financial stability, banking regulation, Basel Accords, Cronbach's alpha

1 Introduction: Impact of bank regulation on stability

Efficiency and stability of the financial sector affect the economy. The institutional environment of a country limits banks' opportunities and thus affects banking sector performance, the economy, and welfare. Banks support the growth of the economy by allocating money to where it is needed most. Or, bank managers and corrupt officials use their influence to support a small privileged group. These two alternative views[1] apply to the regulation as well. Regulators – influenced by politically powerful groups – can work in the public interest and foster a well functioning system or in a private interest for the gain of a few.[2]

The Basel capital accords first passed in 1988 provide a best practice approach to bank regulation which is adopted by practically all countries. Basel II introduced the three pillars approach based on minimum capital requirements, supervisory review, and market discipline.[3,4] The concern about the Basel accords is that the approach mainly developed by rich countries[5] does not fit all countries. In fact, under high corruption it can be counterproductive to give the supervisory authorities more rights[6] as the power might be used to extract further private rents.[7]

[1] Stigler, 1971, p. 3
[2] Compare Barth, Caprio, and Levine (2006, chap. 2) for the public and private interest view of regulation.
[3] Basel Committee on Banking Supervision, 2004, paragraph 4
[4] Basel III concentrates on the first pillar, i.e. capital requirements (Basel Committee on Banking Supervision, 2010, paragraph 1).
[5] An indication is the share of European countries, 10 out of 27, being member of the Basel Committee on Banking Supervision (Basel Committee on Banking Supervision, 2010, footnote 1) where the share of geographical Europe at the world population is well below 10%.
[6] Compare principle 2 and 3 about the scope of supervisory action under "Four key principles of supervisory review" in Basel Committee on Banking Supervision (2004, paragraph 725ff).
[7] Barth et al., 2006, p. 3

The policy question behind my analysis is how countries should implement the Basel accords to reform for the better. Based on the most extensive dataset on worldwide bank regulation, the Bank Regulation and Supervision Surveys of the World Bank, Barth et al. (2006) conducted "the first, comprehensive, cross-country assessment of the impact of bank regulatory and supervisory practices"[8] on banking sector outcomes. For 1999 and 2002 they regress proxies for bank efficiency, bank development, and bank fragility on indices representing the three Basel pillars and conclude (a) that strengthening supervisory review is not advisable as a general rule, and (b) that market discipline should be emphasized more.[9]

In the following analysis I concentrate on the regulatory effect of the three Basel pillars; and I restrict myself to evaluate their impact on bank stability.[10] To investigate the impact of regulation in different institutional environments I separate the country sample along corruption and income.

The rest of the paper is organized as follows. The Bank Regulation and Supervision Survey is discussed in section 2. A cross-section analysis is conducted in section 3 which uses the paper Barth, Caprio, and Levine (2012) as a starting point. It is shown that in contrast to the original the dependent should be modeled in logs and the results cast doubt on the Barth et al. (2006) findings concerning the distinct positive effect of private monitoring. Section 4 incorporates the time dimension. Related to Boudriga, Taktak, and Jellouli (2009) evidence is gained that due to serial correlation in the error terms the authors' results are questionable. Nonperforming loans appear to follow an autocorrelative pattern which I set out to model using dynamic linear panel estimators.

2 Analysis of the Bank Regulation and Supervision Survey

In this section the data is introduced. It is explained why the raw material needs to be aggregated into indices (2.1), how Barth, Caprio, and Levine (2013b) cope with nonavailable items and an evaluation of the indices' validity is given (2.2). The section concludes with descriptive statistics (2.3).

[8] Barth et al., 2006, p. 3

[9] Barth et al., 2006, p. 316

[10] I will use the term *regulation* as a generic description for bank regulation and for bank supervision reviewing the compliance with the rules.

2.1 The data set

In the time frame 1999 through 2011 the World Bank conducted four Bank Regulation and Supervision Surveys, covering about 180 countries. The questions were given to bank regulatory officials and are supposed to represent the formal regulatory situation. The data does not represent how regulation and supervision is actually carried out. This information would be useful to determine the regulatory effect on bank fragility. In this sense the data is second best for the research question of this paper, but nevertheless the best data available.

The questionnaire was given to several officials to cross-check the answers and gaps were filled when information could be attained from other official sources.[11] Due to the lengthy process the surveys do not represent a specific moment, but a time period. Table 1 summarizes meta data on the surveys.

Table 1: Bank Regulation and Supervision survey overview

Survey	Conducted[1]	Representing[2]	Published[3]	Nr. of Countries
I	1999	1999	2001	119-170
II	2003	End 2002	2003	144-168
III	2007	2005-06	2007	128-173
IV	2011	2011-12	2013	116-143

The number of countries covered is given as complete cases (lower bound) and cases where at least one answer was given (upper bound) based on the "All Average Scaled Index" considering the variables used in Barth et al. (2012). [1]Barth et al., 2013b, [2]Barth et al., 2012, p. 2, [3]The World Bank, 2013b.

The World Bank posts the raw survey data of which about 60% are binary questions that are to be answered by "Yes" or "No", about 35% ask for numerical indicators like the share of government-owned banks, and the rest demands to choose between a set of alternatives.[12] The main survey authors Barth et al. use the survey raw data and improve on each of the four surveys resolving inconsistencies and filling gaps. What is more, they aggregate the data to make it directly usable in research.[13] The aggregation is important because single binary questions which capture a sub-aspect of e.g. capital regulatory stringency would not be useful as such to explain financial sector differences like bank fragility.[14] Based on their aggregation they conduct own research, e.g. Barth

[11] Barth, Caprio, & Levine, 2013a, p. 1f
[12] The World Bank, 2013b
[13] Barth et al., 2013a, p. 3ff
[14] To investigate the effect of the three Basel pillars on bank stability one might use many of the binary variables. Following Barth et al. (2013b) this would be 36 variables for the pillars (compare table 2 below). Both the interpretation and the degrees of freedom in the estimation would be a concern.

et al. (2012) as well as other authors like Boudriga et al. (2009). I attempt to reproduce both papers later on.

2.2 Are indices reliable and represent a common factor?

Investigating the impact of regulation on bank fragility relies on reasonably constructed indices. Therefore, the index construction should be reviewed. Barth et al. clearly refer to obstacles and limitations in building proper indices, referring to the process as "[t]he Art and Science of Forming Indices".[15] Even so the authors do not explain and justify their choices. Importantly, the decision which questions enter an index is based solely on theory and the authors' judgment. There is no confirmation based on what the data has to say.[16] Rather, the authors state that the grouping and weighting is not unique and encourage researchers to build their own indices.[17]

Methods to build indices At the core of their approach is to take the sum over the questions entering an index. Mostly the questions are asked such that the answer "Yes" stands for more regulatory stringency and is coded with a 1. One important aspect is how to cope with missing information. Barth et al. provide two different approaches. There is an "All Index" which gives an index value only when the answers to all underlying questions are provided and an "All Average Scaled Index" which allows for missing values. The second index is calculated as the mean of the available items[18] where the index value is only given when at least 50% of the underlying questions were answered and more than two questions enter. Based on this methodology and concentrating on the variables used in re-estimating Barth et al. (2012) there are, varying by survey, 119 to 144 complete cases (shown in table 1). For the "All Index" there are only 47 to 87 complete answers. Setting a threshold for the least available item share is reasonable because uncertainty is associated with this procedure. In the extreme, one answer would decide over the index value summarizing many questions. However, the threshold value of 50% might not be optimal and is, again, not discussed by the authors.

Concentration on variables of interest In the rest of the analysis I concentrate on the bank regulatory variables used in the reproduction of Barth et al. (2012) and Boudriga et al. (2009) in section 3 and 4. Table 2 provides the

[15] Barth et al., 2013a, p. 10

[16] At least I don't see a statistical approach mentioned in Barth et al. (2006), Barth et al. (2013a), or the other publications of these authors mentioned in the references.

[17] Barth et al., 2006, p. 80f

[18] The value is then multiplied by the number of questions entering the index such that both indexation methods yield the same measurement scale.

index definitions and paraphrases the questions the indices consist of. The first three indices represent the Basel pillars capital regulation, official supervision and market discipline. These general aspects of a jurisdiction are represented by ten or more questions. The Capital Regulatory Index tries to capture how limited banks are in using sources as regulatory capital. The first question for this index is in all surveys whether the *Basel I* capital adequacy regime is used and the answer “Yes” is considered as adding to stringency.[19] For 1999 – where *Basel I* still had to be refered to as “Basle guidelines”[20] – this assignment appears fair. However, in 2011 *Basel I* was rather out-dated than strict. This exemplifies the difficulty to balance between continuity over time to increase comparability and adequacy of the questions.

Unidimensionality and reliability The indices presented give a single value summarizing the underlying questions. In the special case of a one-variable index no information is removed (lower part of table 2). When different questions are represented by one value the statistic is misleading except for the case where the questions stand for the same underlying concept. We thus require a common factor behind the questions, or unidimensionality.

Imagine a country’s overall stringency in permitted bank activities is to be assessed. Therefore restrictions in bank insurance and real estate activities are measured. If high restrictions in one activity are associated with high restrictions in the other we do not reject unidimensionality. If, however, either bank insurance activities or real estate activities are allowed the index is not unidimensional. Then the index value as a summary statistic is not informative about the underlying diverging patterns which might influence bank fragility differently. Statistically, a necessary condition for unidimensionality is a positive correlation of the items entering an index. Table 3 gives for all indices used the number of items which exhibit a negative correlation. Notably, there is not a single index which is free of negative correlation between the items. Rather, in many cases a considerable share of items shows a negative association from which I conclude that the indices – or “scales” – offered by Barth et al. are of questionable validity. Using them for analysis introduces a considerable momentum of uncertainty about measuring what we expect to measure. The main hope is that the overall index value is based on a good theory-driven item selection under which the indices are still informative. E.g. when bank insurance activity restrictions are high while real estate restrictions are not we

[19]Barth et al., 2013b, Sheet “Index Overview”, No. IV.I
[20]Barth et al., 2013b, Sheet “Index Overview”, question 3.1.1

Table 2: Bank Regulation and Supervision survey variable definitions

Index (range)	Definition	Questions
Capital Regulatory Index (0-10)	Whether capital requirements reflect risk, market value losses reduce capital adequacy and to which scope sources are accepted to initially capitalize a bank.	Is Basel I used? What fraction of revaluation gains is allowed as part of capital? Are unrealized losses in fair valued exposures deducted from regulatory capital? Are sources of funds verified by authorities? Can the initial disbursement or subsequent injections of capital be done with assets other than cash or government securities (Yes = 0)?
Private Monitoring Index (0-12)	Measures the incentives and the ability for the private monitoring of firms.	Do banks disclose off-balance sheet items to the public? Are bank regulators/supervisors required to make public formal enforcement actions, which include cease and desist orders and written agreements between a bank regulatory/supervisory body and a banking organization? To what extent counts subordinated debt as part of Tier 1/Tier 2 capital?
Official Supervisory Power (0-14)	Whether the supervisory authorities have the authority to take specific actions to prevent and correct problems.	Do banks disclose off-balance sheet items to supervisors? Can the supervisory agency force banks (a) "to constitute provisions to cover actual or potential losses"? (b) "to reduce or suspend dividends to shareholders"? (c) "to reduce or suspend bonuses and other remuneration to bank directors and managers"? Can the supervisory agency force banks to change its internal organizational structure?
Entry into Banking Requirements (0-8)	Whether various types of legal submissions are required to obtain a banking license.	What is legally required before a banking license can be obtained? Draft bylaws, market/business strategy, financial projections, experience of future Board directors, source of funds to be used as capital.
Overall Restrictions on Banking Activities (3-12)	The extent to which banks may engage in securities, insurance and real estate activities.	For the three activity categories a ranking demanded: Can banks directly and fully engage in activities (stringency = 1) up to activities are not allowed in either banks or subsidiaries (stringency = 4).
Variable	range	Definition
Government-Owned Banks	0%-100%	Percentage of banking system's assets in banks that are 50% or more government owned.
Foreign-Owned Banks	0%-100%	The extent to which the banking system's assets are foreign owned.
Bank Concentration (Assets)	0%-100%	The degree of concentration of assets in the 5 largest banks.

Bank Regulation and Supervision Survey indices used in Barth et al. (2012) and Boudriga et al. (2009) are shown. Higher values stand for higher stringency. The indices are grouped into those consisting of multiple questions (upper part) and those out of one variable (lower part). The questions are paraphrased based on surveys I-IV and not exhaustive. Range, definitions, and questions are taken from Barth et al. (2013b) and Barth et al. (2013a, Table 5).

have to assume that in a country with the opposite pattern yielding the same index value the effect on bank fragility is the same to a large extent.

Table 3: Unidimensionality and reliability of the indices

Index (items)	Survey	cor< 0 (cor)	MIC	P_α	α	Countries
Overall Capital Stringency (7)	I	7 (21)	0.12	0.06	0.48	65
	II	10 (21)	0.08	0.06	0.38	73
	III	4 (15)	0.21	0.07	0.65	69
	IV	3 (6)	0.02	0.05	0.13	84
Private Monitoring Index (12)	I	17 (36)	0.03	0.12	0.07	19
	II	17 (36)	0.04	0.09	0.10	68
	III	12 (36)	0.05	0.09	0.12	73
	IV	15 (28)	0.00	0.13	-0.01	90
Official Supervisory Power (34)	I	28 (91)	0.12	0.01	0.82	104
	II	61 (276)	0.12	0.01	0.82	109
	III	68 (253)	0.08	0.01	0.74	119
	IV	24 (66)	0.08	0.01	0.75	121
Entry into Banking Requirements (8)	I	7 (28)	0.16	0.03	0.60	155
	II	0 (28)	0.26	0.02	0.74	159
	III	7 (28)	0.23	0.04	0.70	169
	IV	12 (15)	0.02	0.02	0.14	140
Overall Restrictions on Banking Activities (3)	I	0 (3)	0.31	0.10	0.57	126
	II	0 (3)	0.25	0.08	0.50	151
	III	1 (3)	0.10	0.13	0.24	138
	IV	0 (3)	0.25	0.08	0.50	132
Government-Owned Banks (1)			A single question.			
Foreign-Owned Banks (1)			A single question.			
Bank Concentration, Assets (1)			A single question.			

Measures of internal consistency and unidimensionality per index and survey. All available items are used to calculate the statistics.
cor< 0 (cor) – number of negative item-correlations and in brackets the number of item-correlations in a triangle of the correlation matrix, MIC – mean interitem correlation, P_α – precision of alpha, α – Cronbach's α, Countries – Number of countries for which all index items were answered in that survey. For the remaining variation in number of correlations compare the note concerning incalculable correlations on page 9.

In the rest of this subsection I exploit finer statistical tools for index appropriateness accompanied by the warning that their validity decreases with a higher share of negatively correlated items. The mean interitem correlation (MIC) tries to capture the extent to which items measure the same.[21] To get the MIC one can calculate the mean of all entries in the lower triangle of a correlation

[21] Bühner, 2004, p. 123

matrix for the variables entering an index.[22] I expect the indices to represent a metric latent variable behind the underlying questions. Thus, I use the Bravais-Pearson correlation to calculate the MIC.[23] A correlation of .5 might be interpreted as no clear relation. Therefore a MIC of .5 might serve as a crude tool to assess unidimensionality.

The dispersion of correlations is valuable information not considered in the MIC. A small dispersion of the correlations supports the conclusion of a unidimensional index. The precision of α,

$$P_\alpha = \frac{\sigma_{MIC}}{\sqrt{(1/2 * N * (N-1)) - 1}}$$

is a measure for this, where N is the number of items which enter an index and σ_{MIC} is the standard deviation of the item intercorrelations.[24] P_α increases with intercorrelation dispersion and might thus be called "imprecision of α". As a rule of thumb, Bühner (2004, p. 123) proposes a threshold of $P_\alpha < 0.01$ as supportive for unidimensionality. Table 3 gives both statistics for the used indices.[25] The MIC are all below .5 and P_α is often above 0.01 which indicates a severe problem with multidimensionality in all five indices in the upper part of table 3.

Cronbach's α is often used in the context of evaluating indices. Related to the *MIC* it can be written as (Carmines & Zeller, 1979, p. 44)

$$\alpha = \frac{N * MIC}{1 + (N-1) * MIC}.$$

Essentially the MIC is corrected by the number of items N. It can be interpreted as an upper bound for the unidimensionality of the data.[26] However, it is unclear how far off the true value the measure is.[27] I present Cronbach's α in table 3. When following Cortina (1993, p. 102) a value of 0.75 should be exceeded to be "acceptable" regardless of N. Overall, a low value is a sign that there is a problem – and most of the values are below 0.75. A high value, on

[22] Carmines & Zeller, 1979, p. 44f
[23] Calculating the MIC based on Spearman's rank correlation coefficient does not significantly change the picture. One reason for this is that for 0/1-coded variables the Pearson and Spearman correlation are identical.
[24] Cortina, 1993, p. 100
[25] The values are calculated based on all available information. The reliability can increase when the share of non-available items in an index is limited. I discuss below that the allowed NA share has a negligible influence on Cronbach's α.
[26] Cronbach, 1951, p. 320f
[27] Cortina (1993, p. 101) alludes to the assumption of "tau-equivalence" of Cronbach's α which can be problematic, Bühner (2004, p. 122f) describes how the measure can be biased from adding items which are not useful and that Cronbach's α can be outside the range 0 to 1.

the other hand, does not show the absence of a problem and thus Cronbach's α is to be used with caution.[28]

A necessary condition for an index to be useful is its reliability. According to Bühner (2004, p. 121f) Cronbach's α gives a lower bound for an index' reliability. Thus, a high α gives unambigous information about high reliability but an unclear statement about unidimensionality. When the index items are split in halves, more correlation between the parts is a sign for higher reliability. Cronbach's α is numerically identical to the mean of all possible split-half correlations.[29,30] Based on α, the official supervisory power index has the highest reliability which is rather stable over the surveys. For some indices reliability appears clearly nonsatisfying. Even under the assumption that the indices are useful without being unidimensional low reliability is an important concern about the presented analysis based on Barth et al. (2013b).

Some of the correlations for table 3 were incalculable. E.g. for overall capital stringency and survey III the question "Is the minimum capital-asset ratio requirement risk weighted in line with the 1988 Basle guidelines?"[31] is answered with "No" only by Nigeria and Venezuela. However, both countries gave no answer in a second question for the index. Computing the correlation for this pair is not possible because the standard deviation in the first question is zero. In such cases I dropped the variables lacking relevant variation to calculate the measures in table 3.

Varying the share of NA items I noted above that Barth et al. (2013b) use a threshold of 50% as the smallest share of available items allowed when constructing their "All Average Scaled Index". The threshold is not discussed by the authors. I vary this threshold by 10% steps and control for the change in Cronbach's α. For official supervisory power the largest difference is obtained for survey IV. Allowing a share of non-available (NA) items of at most 0% ("All Index"), 10%, and 20% yields a Cronbach's α of 0.71, 0.73, and .75, respectively.

[28] Another approach to test unidimensionality is factor analysis. Conducting an exploratory factor analysis using principal components based on correlations, the Kaiser criterion proposes for official supervisory power in all surveys at least 5 factors. This is evidence against the appropriateness of assuming one underlying latent variable (a factor). Following the objection that the Kaiser criterion might find too many relevant factors I use Horn's parallel analysis in which still at least 3 factors are proposed (Bortz & Schuster, 2010, p. 415f).

[29] Cronbach, 1951, p. 302ff

[30] Our measures are collected at one point in time and jugded once. When e.g. repeated measurements are available one might measure test-retest reliability. Here we are limited in our choice and test reliability as internal consistency (Cohen, Cohen, West, & Aiken, 2003, p. 129f).

[31] "(...) with the 1988 Basle guidelines?" is taken from Barth et al. (2013b, Index Overview, question 3.1.1). However, in the Excel file it remains unclear which risk ratio is related to. This information is taken from the list of guide question in Barth et al. (2006, p. 337).

Increasing the allowed NA share further leaves the α unchanged. Notably, there is no clear pattern over the different indices and surveys that a higher or lower NA share allowed is associated with a higher α. Additionally, the variation in α appears negligible. In conclusion, a threshold of at last 50% NA items appears to be a reasonable compromise between using the collected information and treating a small amount of answers as representative for a hole index. In this regard my statistical analysis supports the indexation by Barth et al. (2013b).

2.3 Descriptive statistics

Figure 1 gives an impression of the Barth et al. (2013b) dataset coverage. Countries marked in black are contained in the most recent surveys III and IV while those in grey are only contained in survey III.[32] The impression is gained that many African countries left the survey. Though, recognizing that the grey coloured block of central African countries is identical to the members of a financial union[33] it appears likely that the block left due to a common decision. I conclude that there is no particular geographical concentration of the countries leaving the survey. Thus there is some evidence to assume random missingness in the sampled countries which is important for the statistical analysis.

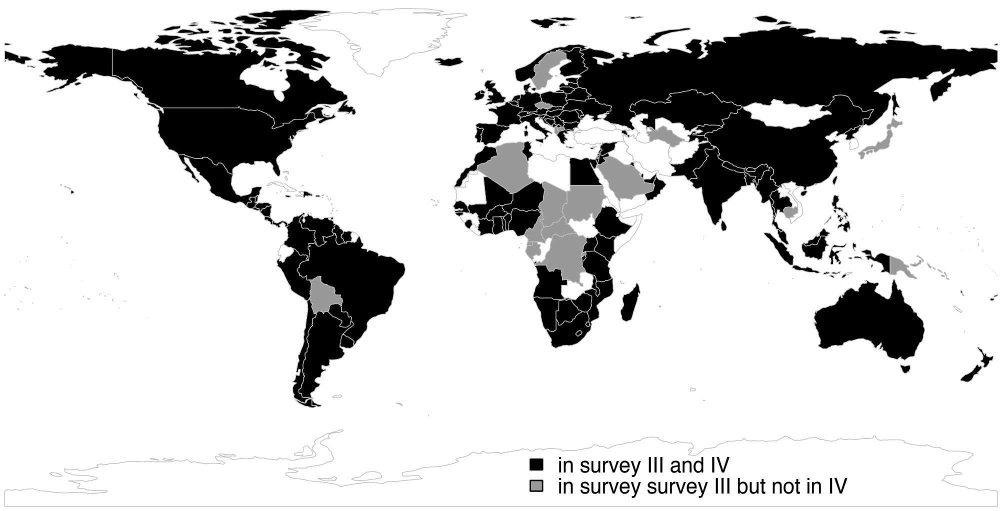


Figure 1: Coverage of survey III and IV on a world map

Comprehensive coverage of survey III, some fewer countries in survey IV.

[32] Coverage is measured based on countries giving at least one answer in the "All Average Scaled Index". Survey III contains the highest number of countries. From survey III to survey IV Ecuador, Gambia, Iraq, Turkey, and Yemen were added.

[33] The grey block represents the Central African Economic and Monetary Community (CAEMC) consisting of Cameroon, Central African Republic, Chad, Congo, Equatorial Guinea, and Gabon (Barth et al., 2013b, "Groups with uniform bank regulations and supervisory practices").

Accepting a random country coverage, I am interested in the answer pattern of the sampled countries. One dimension is whether the missingness by question appears to be random; the share of countries for which the indices are available ranges between 66-91% with a standard deviation of about about 10%. I consider the dispersion as not too large. To obtain these values I look at all four surveys, use the "All Average Scaled Index" version, and restrict on the indices used to reproduce Barth et al. (2012) and Boudriga et al. (2009). For the "All Index" – where a missing value in a items leads to a NA for the index – only 47-87% are available with a slightly greater dispersion. Based on these numbers and the findings varying the allowed NA share I use the "All Average Scaled Index" throughout the subsequent analysis.

Another dimension is whether the missingness related to answers per country appears to be random. Figure 2 shows that 35 countries gave answers such that 1-4 index values can be calculated in 1999 (survey I). However, zero countries answered in such a way in 2011 (survey IV). Questions added to survey IV due to the financial crisis might have triggered either survey fatigue or the motivation to answer many questions. In this indirect way the change in other survey questions might affect the indices used in the analysis which are build on as tantamount questions as possible to allow comparability over the horizon of over 10 years. Survey IV stands out again as the study answered by the least countries. These issues might flaw the identification of patterns; I assume that this is not the case.

The standard deviations of the three indices representing the Basel pillars are nearly unchanged comparing 1999 and 2011. The mean stringency of capital regulation clearly and of private monitoring marginally increased while mean supervisory power marginally decreased. Underlying the values is a considerable dispersion. Capital stringency (index range is 0-10) was decreased over the 12 years by United Kingdom, Austria, and Mexico by 5 while Venezuela, Turkey, and Bangladesh increased it by 6. In Kazakhstan official supervisory power (index range is 0-14) was decreased by about 8 while Italy increased it by 7. Private monitoring (index range is 0-12) was decreased by 2 in Portugal but increased by 4 in India and France.[34]

[34]The information is obtained from Barth et al., 2013a, figures 9-11. In the paper much more descriptive statistics about the studies are available. Germany increased capital regulations by 2, both official supervisory power and private monitoring by 1, and entry into banking requirements by 4 (only Chile and Finland increased the stringency more).

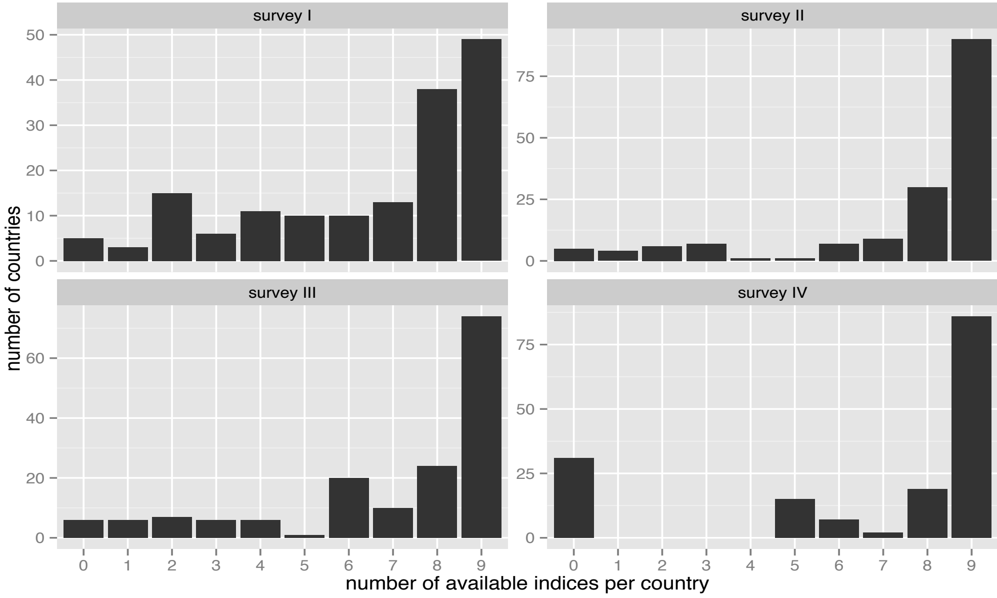


Figure 2: Number of answers in Bank Regulation and Supervision Survey I-IV
Nine of the indices provided by Barth et al. (2013b) are used in the analysis of bank fragility. The distribution of available indices by survey is shown.

3 Explaining nonperforming loans (NPL)

This section provides a cross-section analysis of the Basel pillars' impact on nonperforming loans (NPL). The Bank Regulation and Supervision surveys approximately represent certain years and therefore this approach is natural. First, I attempt to reproduce the results of Barth et al. (2012); this is shown not to be possible due to data issues. The analysis in subsection 3.1 can be seen as a robustness check; while significance differs somewhat the results are in no case contradictory. Subsection 3.2 shows that the dependent should be modeled in logs. Doing so casts doubt on the Barth et al. (2006) finding of a distinctly positive effect of the third Basel pillar, private monitoring. Subsection 3.3 shows that the results are robust to removing outliers while 3.4 introduces an approach to reduce endogeneity in the cross-section setting. Subsection 3.5 investigates differences along income and corruption.

3.1 Reproduction of Barth, Caprio, and Levine (2012)

Barth et al. (2012) investigate the determinants of NPL as a proxy for bank fragility. Their cross-section approach for 1999 and 2011 suffers from a small amount of observations which reduces the chance to obtain significant results. The upside of this approach is that problems due to incorporating the time dimension are prevented. In their regressions Barth et al. assume that the impact of regulation is the same in all countries,

$$\begin{aligned} \mathrm{NPL}_i &= \alpha + \mathrm{regulation}_i'\beta + \mathrm{control}_i'\gamma + u_i \qquad (1) \\ i &= 1, \ldots, N, \text{ for } 1999, 2011, \; u_i \sim \mathrm{N}(0, \sigma_{\mathrm{i}}^2) \end{aligned}$$

where β is a vector of dimension $n_1 \times 1$ and γ of $n_2 \times 1$ where n_1 and n_2 stand for the number of regulatory variables and control variables, respectively. The vectors $regulation_i'$ and $control_i'$ are correspondingly of dimension $1 \times n_1$ and $1 \times n_2$. The error is assumed to be normally distributed with a mean of zero and different variances, i.e. heteroscedasticity.

Definitions and sources of the variables used are shown in table 4. As regulatory and supervisory variables the authors include the Basel accord pillars – capital regulation, official supervision, and market discipline, the last one also called private monitoring – as well as entry requirements and activity restrictions. Barth et al. skip all non-complete cases: "(...) for countries with missing values, [we] had to drop that observation (...)".[35] The regression output in Barth et al. (2012, p. 17) shows 68 observations entering the ordinary least squares estimation for 1999. Accordingly, at least 68 observations of the regulatory variables without any missing values have to be available.

However, the dataset Barth et al. (2013b) contains only 44 complete cases for the 1999 regulatory variables entering regression 1. I cannot explain this difference; however, I observe that 109 complete cases are available for the "All Average Scaled Index" and I assume that the authors mistook "unscaled" and "scaled" in their description. I decide to use the richer version of the data which gives index values even when some items are missing.[36]

The amount and scope of control variables in regression 1 is moderate. Barth et al. use only the countries legal origin.[37] NPL are clearly influenced by the economic situation and therefore a time-invariant control variable like legal

[35] Barth et al., 2012, p. 15

[36] The working paper was published in December 2012. I received the 2013 dataset from the homepage of one of the authors (Barth et al., 2013b). Data updates in the 2013 version of the data, e.g. due to going back to the authorities and correcting entries, would have increased the complete cases. I assume that the large discrepancies cannot be explained by information that had to be dropped for some reason. The legal origin variables introduced next are available for all countries of survey I and IV and cannot be the reason for the difference.

[37] The source for the legal origin dummies is not stated in the working paper Barth et al. (2012). Due to the similar approach I assume that the same source as in their earlier work Barth et al. (2006, p. 191f) is used, namely data from La Porta et al. The co-author Shleifer provides the data on his website (La Porta et al., 2008).

Table 4: Variable definitions and data sources for Barth et al. (2012)

Variable	Definition	Original source
Bank nonperforming loans to gross loans (%), abbreviated: NPL	Ratio of defaulting loans (payments of interest and principal past due by 90 days or more) to total gross loans (total value of loan portfolio). The loan amount recorded as nonperforming includes the gross value of the loan as recorded on the balance sheet, not just the amount that is overdue. Note that due to differences in national accounting, taxation, and supervisory regimes, these data are not strictly comparable across countries.	IMF (2008) for 2002-04, IMF (2011) for 2005-07, IMF (2013) for 2008-13
Capital Regulatory Index	Whether capital requirements reflect risk, market value losses reduce capital adequacy and to which scope sources are accepted to initially capitalize a bank.	Barth et al. (2013b)
Private Monitoring Index	Measures whether there are incentives/ability for the private monitoring of firms, with higher values indicating more private monitoring.	Barth et al. (2013b)
Official Supervisory Power	Whether the supervisory authorities have the authority to take specific actions to prevent and correct problems.	Barth et al. (2013b)
Entry into Banking Requirements	Whether various types of legal submissions are required to obtain a banking license.	Barth et al. (2013b)
Overall Restrictions on Banking Activities	The extent to which banks may engage in securities, insurance and real estate activities.	Barth et al. (2013b)
Government-Owned Banks	Percentage of banking system's assets in banks that are 50% or more government owned.	Barth et al. (2013b)
Legal origin	Dummies for legal origin English, French, German, and Scandinavian.	La Porta, Silanes, and Shleifer (2008)

origin appears insufficient. In this section I stick to the authors' approach while section 4 will use a more comprehensive set of control variables.

Capturing bank fragility with NPL Barth et al. use as dependent variable NPL as a share of total assets. The share increases when NPL increase which puts pressure on the banks' equity and is thus a good indicator for bank fragility. Nonetheless, if total assets increase and are used for speculation in less regulated fields, NPL as a share of total assets can go down although risk and bank fragility increase. This calls for a more confined concept. I will use NPL as a *share of total loans*. Additionally, NPL as a share of total loans is publicly available while the broader concept used by Barth et al. is not.[38] NPL as a share of total loans is provided by the International Monetary Fund and is available for 77 and 92 countries in 1999 and 2011, respectively.[39] The NPL variable thus provides the least observations in regression 1 and limits the estimation's scope. The NPL data comes with a warning: "Due to differences in consolidation methods, national accounting, taxation, and supervisory regimes, data are not strictly comparable across countries." (IMF, 2013). I translate this additional uncertainty into a need for strong empirical results.

Correlations The common variation of the variables entering regression 1 are shown in table 5. NPL to total loans and the share of government-owned banks represent at least an interval scale of measurement. The indices summing up the individual yes/no-questions are assumed to represent a metric concept. Accordingly, the Bravais-Pearson correlation is used. Spearman's rank correlation as alternative does, however, not qualitatively change the picture.

Higher values represent higher stringency. In 1999 both capital regulatory stringency and private monitoring are associated with lower bank fragility measured by NPL. In contrast, the second Basel pillar supervisory power is associated with higher bank fragility. In 2011 the common variation between the Basel pillars and NPL is basically lost. Thus, we expect to find a clear impact of the regulation countries adopt and the countries' NPL level in 1999 while we do not expect a strong impact in 2011. In the aftermath of the financial crisis in 2011 NPL are probably too much affected by contagion, i.e. the financial turmoil which started in the US and spread to other countries regardless of the regulation implemented.

The share of government-owned banks will be added to equation 1 in a second step. In 1999 the correlation of the share of government banks with private monitoring is $-.37$. This represents that countries with a higher share of government banks – which basically face the same default risk – provide fewer incentives for the market monitoring of risk taking. Because more government banks are (a) associated with more fragility and (b) associated with less private

[38] NPL as a share of total assets is not available from the World Bank and I was not able to receive it elsewhere in the necessary scope of a variety of the worlds' countries and the time frame 1999-2011.

[39] The NPL as a share of total loans for 1999-2011 had to be collected from three sources. From IMF (2008) to IMF (2011) I noted no differences in overlapping years. However, from IMF (2011) to IMF (2013) some overlapping country series exhibit clearly different numbers. The IMF Data Dissemination and Client Services Team answered to my inquiry that responsibilities changed within the relevant period inducing a change in methodology in April 2011. Whereever data overlaps I use the values from the newest source available.

Table 5: Correlations of variables in Barth et al. (2012)

(a) For 1999 / survey I

	NPL	CapReg	PrivMon	SupPow	Entry	ActRes	GvtBk
NPL	1						
CapReg	−0.23	1					
PrivMon	−0.29	0.13	1				
SupPow	0.28	−0.04	0.19	1			
Entry	0.12	0.16	−0.11	0.21	1		
ActRes	0.38	0.04	−0.09	0.04	0.06	1	
GvtBk	0.33	−0.06	−0.37	−0.14	−0.13	0.30	1

(b) For 2011 / survey IV

	NPL	CapReg	PrivMon	SupPow	Entry	ActRes	GvtBk
NPL	1						
CapReg	0.02	1					
PrivMon	0.03	0.15	1				
SupPow	−0.01	0.11	0.11	1			
Entry	−0.03	0.07	−0.08	−0.01	1		
ActRes	−0.17	0.32	0.06	0.19	0.12	1	
GvtBk	0.12	0.04	0.01	−0.07	0.02	0.02	1

The common variation of NPL and the three Basel pillars reduced significantly after financial crisis starting in 2007. Higher values of the bank regulatory and supervisory variables stand for more stringent regulation. The "All Average Scaled Index" of Barth et al. (2013b) is used. CapReg is capital regulation, PrivMon is private monitoring, SupPow is supervisory power, Entry are Bank entry requirements, ActRes are overall restrictions on banking activities, GvtBk is the share of government-owned banks.

monitoring linked to more fragility I expect the introduced variable to reduce the significance of private monitoring.

Estimation The most relevant difference compared to Barth et al. (2012) is the dependent NPL to total loans versus their variable NPL to total assets. In this I see the main reason for the different number of observations entering the regressions in table 6. Model 1 uses 62 observations while the original is based on 68.[40] First, I run the regressions without heteroscedasticity-consistent standard errors and investigate with White's general test to what extent the variation in the residual variance is a problem. The Null of the test is H_0: $\sigma_i^2 = \sigma^2$ for all i, i.e. homoscedasticity. The squared regression residual is explained in a linear regression by the explanatory variables and all of its squares and cross products plus an intercept. The underlying assumption is that the functional terms are

[40] Adding government banks reduces the observations in the original paper more drastically than in my regressions. As I almost surely have the same information on the share of government banks I think this corresponds to a different country sample selected due to the distinct dependent variable.

sufficient to identify any heteroscedasticity. The kind of variation in the errors does not have to be specified.[41] A downside of the test is that the multiplicity of explanatory variables in the auxiliary regression reduces degrees of freedom and can even render the estimation infeasible.

From 62 observations 36 degrees of freedom remain in the case of the first regression in 1999 where I obtain a multiple R^2 of 41% explaining $\hat{u}_i^2$. The test statistic is χ^2-distributed with $nR^2 \sim \chi^2_{K(K+1)/2}$ where K is the number of regressors including the intercept such that $n * R^2 = 62 * .41 = 25.42 < \chi^2_{9(9+1)/2} = 51.00$ such that the Null of homoscedasticity is not rejected for the baseline regression in 1999.[42,43] Repeating the exercise for 2011, i.e. model 3 in table 6, yields a p-value < 0.001 indicating heteroskedasticity in the data.[44]

Table 6: OLS regressions re-estimating Barth et al. (2012)

	for 1999		for 2011	
	(1)	(2)	(3)	(4)
Capital Regulation	-0.015^{**}	-0.014^{*}	0.003	0.003
	(0.007)	(0.007)	(0.004)	(0.004)
Private Monitoring	-0.021^{**}	-0.011	0.001	0.002
	(0.009)	(0.012)	(0.005)	(0.005)
Official Supervisory Power	0.009^{*}	0.007	0.002	0.002
	(0.005)	(0.006)	(0.004)	(0.005)
Entry Requirements	0.004	0.007	0.012	0.006
Restrictions on Activities	0.017^{***}	0.014^{**}	-0.006^{*}	-0.007^{*}
Government-Owned Banks		0.001		0.0004
Observations	62	56	76	69
Adjusted R^2	0.293	0.264	0.036	0.045
F Statistic	4.166^{***}	3.191^{***}	1.348	1.353

*p<0.1; **p<0.05; ***p<0.01
Dependent variable: nonperforming loans to gross loans. Legal origin variables and the constant are suppressed.

Heteroscedasticity might not be present in all regressions in the reproduction of Barth et al. (2012). Still, the application of heteroscedasticity-consistent standard errors has a minor impact on the inference of a regression without heteroscedasticity. Hence, I follow the original to use a heteroscedasticity consistent (HC) covariance matrix estimation for all regressions. Barth et al. do

[41] Greene, 2008, p. 165f
[42] White, 1980, p. 824f, and Wang, 2014
[43] The probability for the Null (p-value) is 90.1%.
[44] The example in Greene (2008, p. 167) indicates that White's test might be quite conservative. For 1999 homoscedasticity might not be rejected even so it is present. However, applying the Breusch-Pagan/Godfrey LM test using Koenker's version not sensitive to violations of normality – as proposed in Greene, 2008, p. 166 – yields the same test decisions for the two equations.

not state which of the HC estimators they use. HC3 is a variant which builds on White (1980) and tries to additionally correct for a bias due to small samples. Following the Monte Carlo simulations in Long and Ervin (2000) indicating HC3 to be the best of the alternatives I use HC3 in the cross-section analysis.

The regressions in table 6 confirm the findings from the correlations and the general pattern fits the Barth et al. (2012) results. In 1999 capital regulatory stringency on average significantly decreases bank fragility. More private monitoring has a similar stabilizing effect but the result is less robust as it is lost when controlling for government banks. Theory tells us that the impact of official supervisory power is moderated by the rule of law and corruption in a country. The only marginal and not robust effect in 1999 is therefore not surprising.[45] In 2011 (model 3 and 4) the overall F-test cannot reject $H_0 : \beta = \gamma = 0$ (compare equation 1) at a level of uncertainty of 10%. This is evidence that after the financial crisis starting in 2007 bank fragility is driven to a large extent by contagion and less by countries' regulatory decisions.

Beyond a statistical significance we are interested in the effective size of an effect. Observing in model 1 of table 6 a numerically more extreme value for private monitoring than for capital regulation, in absolute terms, cannot be directly evaluated in terms of effect size. The reason lies in the different measurement scales used, e.g. different ranges for the Basel pillars (compare table 2). Among the Basel pillars I expect capital regulation to have the strongest impact on NPL as its coefficient is higher than those of the other Basel pillars which have about the same range. Using standardized regression coefficients confirms the assessment.[46]

3.2 Can we trust the estimation results?

Multicollinearity is present if one of the explanatory variables can mostly be replaced by information in the other regressors. This can have serious conse-

[45] The impact of activity restrictions on fragility is not clear in theory (Barth et al., 2006) and here found to be positive while not significant in Barth et al. (2012). I don't consider this further as the research question is the impact of the Basel pillars.

[46] Standardized regression coefficients are based on the estimated coefficient and "standardize" by multiplying with sd_i/sd_{NPL}, i.e. the standard deviation of the evaluated variable divided by the standard deviation of the dependent (Bring, 1994, p. 210). Bring, 1994, p. 211 criticizes that using both standard deviations and coefficients is inconsistent as coefficient estimates assume the other regressors are held constant and thus relate to another population than the standard deviations based on the full sample. Bring proposes to use the variance inflation factor (VIF) introduced above and jugde coefficients as less relevant when their VIF is higher, i.e. when similar information is contained in other regressors. The result described is obtained based on either the standard formula $\hat{\beta}_i * (s_i/s_y)$ or using Brings' proposal $\hat{\beta}_i * (sd_i/\sqrt{VIF_i}) * \sqrt{n-1/n-k}$.

quences for estimation, t-tests, and inference. Multicollinearity cannot be sufficiently identified based on the correlation matrix which shows only collinearity, i.e. the common movement in two variables. I use the Variance Inflation Factor (VIF) as one possible test for the multicorrelation in the independent variables. The VIF of an independent variable is calculated as $V_i = 1/(1 - R^2_{(i)})$ where $R^2_{(i)}$ is obtained as the R^2 regressing x_i on a constant and the remaining regressors.[47] A high R^2 in this auxiliary regression indicates common movement with one or more of the variables and leads to a high V_i value.

In table 6 the regressions contain a dummy variable for English, French and German legal origin while Scandinavian legal origin is omitted to prevent perfect multicollinearity. For all regulatory variables the VIF is below 1.5; values above 5 or 10 are given as rules of thumb indicating a problem. The legal origin dummies give clearly higher VIF up to about 6. As control variables they are included to lessen the omitted variable bias of the parameters of interest. Notably, multicollinearity in control variables does not affect the estimation of the Basel pillars' effect on bank stability.[48] In sum, I reject multicollinearity to harm the identification.

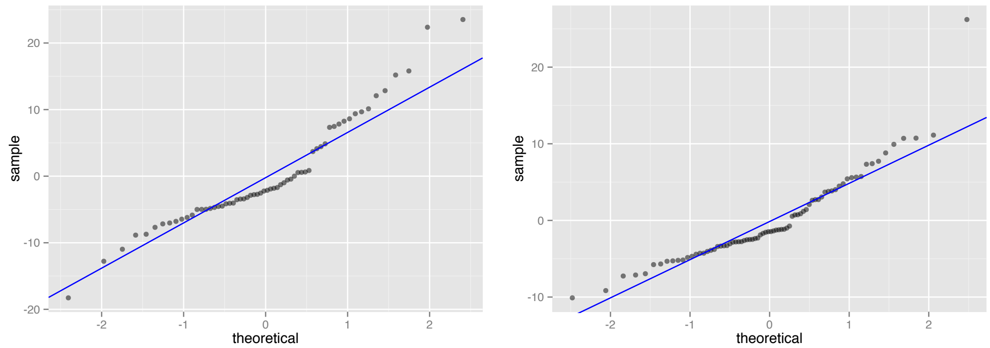
Non-normality In contrast, non-normality of the errors is found to be a severe flaw in the re-estimated regressions. Figure 3 plots quantiles of the OLS residuals against those of the normal distribution. Figure 3a is based on regressions 1 and 3 in table 6, the baseline regressions for 1999 and 2011. Clear deviations particularly in the central part of the distribution are apparent.

I suspect that the strong positive skewness of NPL is an important driver of the model misspecification.[49] I use a log transformation of the dependent variable to reduce the skewness. As all NPL values are positive we can do so. The log transformation attempts to linearize the relation between the independent variables and NPL. A power transformation chosen more specific to the data would yield a better fit but would come at the cost that both interpretation becomes harder and predictive power can be harmed due to overfitting. Figure 3b shows the normal probability plot for residuals based on the dependent log(NPL), other things equal. Now for 1999 the residuals fit much better to normality. For 2011 the fit improves as well, but to a smaller extent; still a big part in the middle of the distribution does not fit the normal distribution. The

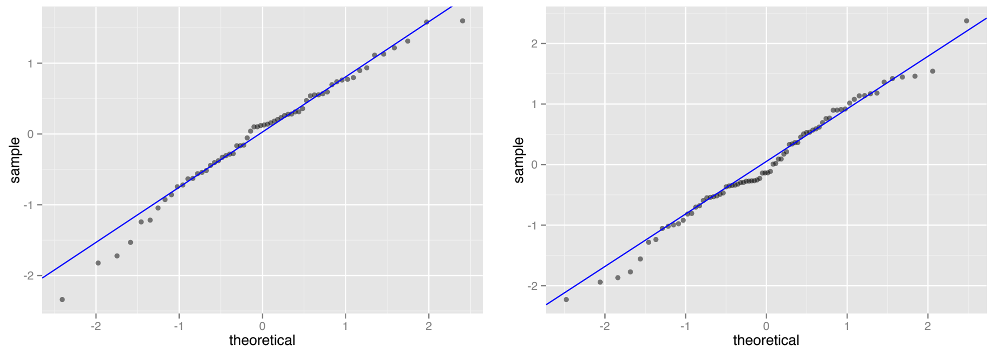
[47] Mela & Kopalle, 2002, p. 676f

[48] Von Auer, 2007, p. 488f

[49] For the variables entering the regressions in Barth et al. the sample skewness attains values of +1.5 and +2.3 for 1999 and 2011, respectively. The values slightly differ by the method used. I employ $\sqrt{n-1} * m^3/(m^2)^{3/2}$ where $m^r = \sum_i^n (x_i - \bar{x})^r$ stands for the sample moments of order r (e.g. Greene, 2008, p. 1020f).



(a) Dependent variable is NPL for 1999 and 2011, respectively



(b) Dependent variable is the *log of* NPL for 1999 and 2011, respectively

Figure 3: Normal probability plots for regressions related to Barth et al. (2012)
Plots are for regression 1 and 3 in table 6 (upper left, upper right) and for regression 1 and 3 in table 7 (lower left, lower right).

improved fit can be shown with a normality test on the regression residuals. The Shapiro-Wilk tests' Null hypothesis of normality is rejected for the errors of all models in table 6 ($p < .10$), i.e. modeling NPL. Yet, modeling *log*(NPL) yields p-values well above 10% confirming the visual better fit of the errors.[50]

The improved specification modeling $log(NPL)$ is shown in table 7. The coefficients in 2011 (models 3 and 4) are again not statistically different from zero based on the overall F-test with $p = .10$ used as threshold. But for 1999 the overall F-test statistic and the adjusted R^2 increase considerably. Capital regulatory stringency increases bank stability with moderate significance in table 6. Modeling *log*(NPL) capital stringency is identified as a highly significant driver of NPL. Under the inclusion of government banks the result is still significant at

[50]Shapiro and Wilk (1965, p. 608) find their test to be quite sensitive to deviations, e.g. to asymmetry. The Shapiro-Wilk test has drawbacks for large sample sizes (Shapiro & Wilk, 1965, p. 610) which is no problem here with below 100 observations.

$\alpha = 5\%$ (model 2 in table 7) while the analog in Barth et al. (2012), modeling NPL, shows an insignificant first Basel pillar.

Table 7: OLS regressions of Barth et al. (2012) with log(NPL)

	for 1999		for 2011	
	(1)	(2)	(3)	(4)
Capital Regulation	-0.201^{***}	-0.182^{**}	0.050	0.048
	(0.070)	(0.076)	(0.067)	(0.063)
Private Monitoring	-0.156	-0.029	0.015	0.022
	(0.094)	(0.128)	(0.085)	(0.079)
Official Supervisory Power	0.018	0.007	0.010	0.0004
	(0.057)	(0.069)	(0.058)	(0.062)
Entry Requirements	0.055	0.117	0.080	-0.087
Restrictions on Activities	0.223^{***}	0.190^{**}	-0.060	-0.076
Government-Owned Banks		0.015^{**}		0.003
Observations	62	56	76	69
Adjusted R^2	0.376	0.407	-0.024	0.021
F Statistic	5.597^{***}	5.196^{***}	0.778	1.162

$^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$
Dependent variable: Log of nonperforming loans to gross loans. Legal origin variables are suppressed.

The results cast doubt on private monitoring as a key driver of bank stability. Modeling log(NPL) the positive effect on bank stability is lost.[51] Private monitoring being an unmitigated good was a result of the book “Rethinking bank regulation” by Barth et al. based on studies I and II in 1999 and 2002.[52] In Barth et al. (2012, p. 17ff) private monitoring turns insignificant in 1999 when government banks are included. This questions the robustness of the positive impact of the 3^{rd} Basel pillar. Nevertheless, when I run the model in equation 1 for 2002 based on the dependent log(NPL)[53] private monitoring becomes significant at the 5% level and capital regulation turns insignificant – both in contrast to the 1999 results. For survey III and 2006 no Basel pillar is significant while the overall F-test rejects the hypothesis that all coefficients on the regressors are zero. In 2011 I found no statistical evidence for any pattern in the baseline model. In this sense from survey I in 1999 till survey IV in 2011 less and less of the variation in NPL can be explained.

I conclude that the patterns are not robust over time. This could either represent the true underlying pattern. Or it is (a) the result of outliers leading to a biased picture which is discussed in the next subsection; (b) due to the contemporaneous nature of the left- and right-hand side variables which is discussed in

[51] The coefficient has a p-value of 0.103 for model 1 in table 7.
[52] Barth et al., 2006, p. 316
[53] To conserve space that regression is only shown below as model 1 in table 8b.

subsection 3.4; (c) due to diverging country patterns not adequately modeled by pooling. E.g. income and corruption might mediate the Basel pillar effect on NPL which is discussed in subsection 3.5.

3.3 Outliers and strong influence

A concern about the regressions is that a small amount of countries with a deviating trend blurs the view on the overall pattern. Then the results would represent a mix of the deviating and the overall trend. We can identify outliers, or unusual observations, using the prediction error, i.e. the residual $u_i = y_i - \hat{y}_i$ for country i. To make the deviation comparable I use the studentized regression residuals to identify countries with a pattern outside the 95% confidence interval of a normal distribution.[54] Figure 4 shows the absolute values of the studentized residuals exemplified for the 1999 baseline regression explaining log(NPL).[55] Based on the approximation $|z| > 2$ the United States, Chile and Luxembourg are outliers in the sense that they don't fit the overall pattern.

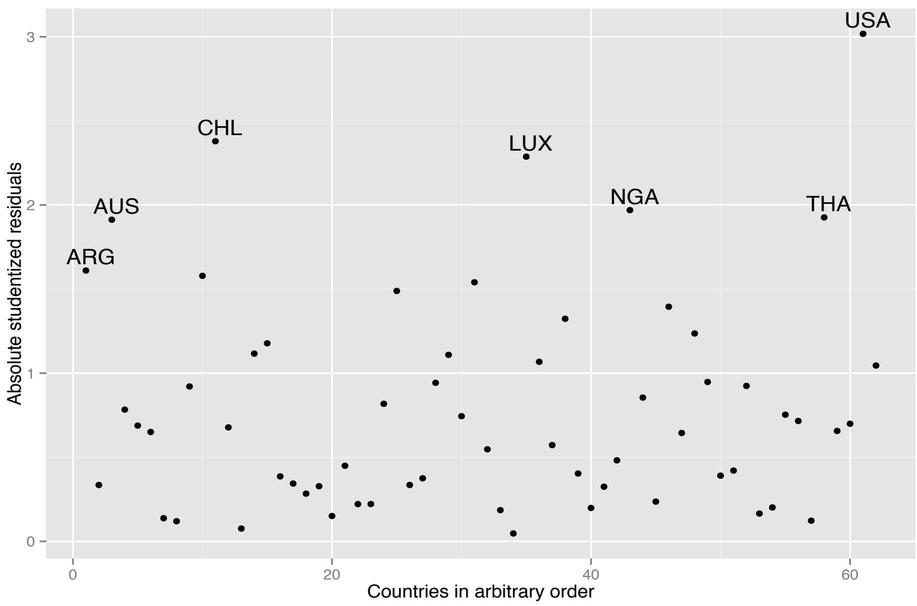


Figure 4: Outliers in OLS regression

Identification of countries with a deviating pattern based on studentized residuals exemplified for model 1 in table 7. Country codes are United States (USA), Chile (CHL), Luxembourg (LUX), Nigeria (NGA), Thailand (THA), Australia (AUS), and Argentinia (ARG).

[54]Studentized residuals are preferred over standardized residuals to identify outliers (Cook & Weisberg, 1982, p. 37). Studentized residuals additionally consider different residual variances which is relevant under heteroscedasticity.

[55]Model 1 in table 7.

Removing outliers has a minor impact For the baseline model and each survey year I separately drop the countries outside ±2 and control for the impact on the coefficients and significance. For 1999 the results of model 1 in table 7 do not qualitatively change; the adjusted R^2 increases by 13 percentage points. Similarly, for 2002, 2006, and 2011 the results on the effect of the Basel pillars on NPL do not change when the outliers are dropped.[56]

Strong influence Residuals are not sufficient to identify influential observations. Imagine a perfect positive linear relationship between NPL as dependent and supervisory power as regressor with many observations around the mean of supervisory power. A country is added with a mean supervisory power value but unusually high NPL. The regression line will not considerably shift because there are many other observations around the mean of the explanatory variable. The new observation produces a huge error $|y_i - \hat{y}_i|$ and is identified based on absolute studentized residuals. Now a country is added with clearly the highest level of supervisory power and an NPL level different from the linear trend. Due to little information at this level of the independent variable the regression line will adapt to the new observation. Therefore a small error results and the observation is not identified by studentized residuals while the estimated coefficient β is considerably affected.[57] Cook's distance proposed by Cook (1977) evaluates the change in $\hat{\beta}$ from dropping an observation and is thus sensitive to influential observations. Dropping the countries which stand out based on Cook's distance does not qualitatively affect the results.[58] Cook's distance is hardly helpful to identify clusters of countries as just single observations are dropped. That said, dropping countries identified by absolute studentized residuals and Cook's distance speaks for the validity of the regression results in table 7.

3.4 Contemporaneous nature

I am concerned with measuring the effect of the Basel pillars implementation on countries' bank fragility. Observing a change in regulation, the subsequent impact on bank fragility should be evaluated. In section 3.1 NPL in 1999 are explained by the regulation in 1999. But regulation might only have an impact on NPL after some time rendering the identification strategy inappropriate.[59]

[56] For 2002 entry into banking requirements turn significant at the 10% level. For 2006 the significance of activity restrictions reduces from the 10% to the 5% level.

[57] Cook, 1977, p. 16f

[58] I drop Chile for 1999 and 2002 which is clearly the most influential country. For 2006 I drop Luxembourg and Canada, for 2011 Iceland and Finland.

[59] Barth et al. write that "the contemporaneous nature of the right- and left-hand side variables raises a caution in the interpretation" (Barth et al., 2012, p. 16).

Measuring both explanatory and explained variables in the same year might induce endogeneity because observing higher NPL can lead to calls for regulatory intervention. Explaining NPL by last year's regulation is a natural approach,

$$\begin{aligned} log(\mathrm{NPL}_{it}) &= \alpha + \mathrm{regulation}'_{i,t-1}\,\beta + \mathrm{control}'_i\,\gamma +\ u_i \\ i &= 1,\ldots,N,\ \text{for}\, 2000,\ 2003,\ 2007,\ u_i \sim \mathrm{N}(0,\sigma_{\mathrm{i}}^2) \end{aligned} \tag{2}$$

where it is assumed that NPL exhibits no longer-term trends. Using more lags of regulation reduces endogeneity but comes at the cost that the regulatory impact might have faded out already. Based on the compromise to use one lag – explaining with the 1999 regulation $log(\mathrm{NPL}_{i,2000})$, with the 2002 regulation $log(\mathrm{NPL}_{i,2003})$ and with the 2006 regulation $log(\mathrm{NPL}_{i,2007})$ – no qualitative difference in the Basel pillar coefficients is found. For 2012 no NPL data is available such that the approach is not possible for survey IV. This robustness is another hint that the estimated cross-section model is appropriate.

Still, endogeneity remains a concern. Countries with a high equilibrium NPL level are more likely to enact tighter regulation. This country-specific effect is not modeled in this section and would thus be captured by the error term which then correlates with the regulatory variables. Section 4 sets out to approach this issue incorporating the time dimension. Before that – and under the assumption that endogeneity does not lead to flawed inference in the cross-section – the next subsection approaches the policy question which countries gain, i.e. can reform for the better, by tightening regulation related to the Basel pillars.

3.5 Modeling country differences

I gained evidence that a valid model to explain NPL in a cross-section approach is obtained. This forms the basis to investigate the distinct impact of regulation on bank stability which theory predicts for different institutional environments. Therefore, I divide the country sample into groups and examine in which sample the Basel pillars have a more profound impact.

The first aspect I control for is income. This is of interest as the Basel rules are mainly produced by the rich countries and might thus be more adequate for these. Income might be accepted as a proxy for development, the dimension along which Boudriga et al. (2009, p. 301) divide their sample.[60] As introduced

[60] Boudriga et al. do not make explicit how they decide which countries are developed and which are developing. Based on that information the results would be more explicitly comparable.

in section 1 corruption will matter – e.g. for more supervisory power having a positive or a negative effect on stability – and is the second aspect I control for.

To investigate the effect of income, a dummy variable is used that takes 1 if the 2012 GNI per capita is equal to or above $12,616 ("high income") and 0 else. This corresponds to the recent World Bank classification.[61] Based on the ranks of the Corruption Perception Index[62] where the first rank stands for the least corrupt country I build a corruption dummy that takes 1 if a country rank in a certain year is worse than the mean rank and 0 else. Hence, corrupt countries are assigned a 1.

Statistically, one might divide the countries into two groups and estimate the coefficients separately, or one includes an interaction effect. In separate regressions one assumes the variance of the errors to be different, with the interaction term to be the same.[63] The separate estimation has the downside to use less information and thus leads to less significance.[64] I use interaction effects where a significant interaction can be directly seen from the statistic on the associated coefficient. I add the interactions of interest separately mainly to prevent a strong decrease in degrees of freedom.[65] The model for the income dummy variable $\mathrm{Inc_i}$ and the interaction with capital regulation can be written as

$$\begin{aligned} log(\mathrm{NPL}_i) &= \alpha + \mathrm{regulation}_i'\,\beta + \mathrm{control}_i'\,\gamma \\ &\quad + \delta\,\mathrm{CapitalRegulation} * \mathrm{Inc}_i + \zeta\,\mathrm{Inc}_i + u_i \qquad (3) \\ i = 1, \ldots, N,&\ \text{ for } 1999,\ 2002,\ 2006,\ 2011,\ u_i \sim \mathrm{N}(0, \sigma_\mathrm{i}^2) \end{aligned}$$

where the vector *regulation* for country i contains the capital regulatory variable in the first position as well. The model can be rewritten as

$$\begin{aligned} log(\mathrm{NPL}_i) &= (\alpha + \zeta\,\mathrm{Inc}_i) + (\beta_1 + \delta\,\mathrm{Inc}_i) * \mathrm{CapitalRegulation}_i \\ &\quad + \mathrm{regulation}_{-j,i}'\,\beta_{-j} + \mathrm{control}_i'\,\gamma + u_i \end{aligned}$$

where the index $-j$ represents the remaining regulatory variables collected in $regulation_{-j}$. From this formulation it is most easy to see that for the set of countries with a per person income below the threshold ($\mathrm{Inc}_i = 0$) we estimate

[61] The World Bank, 2014a.
[62] Transparency International (2013) in the time frame 1999-2011 is the original source while the data summarized in one file was obtained from DICE Database (2014).
[63] Von Auer, 2007, p. 320
[64] Jaccard & Turrisi, 2003, p. 36
[65] In table 8 the models contain up to 9 variables plus an intercept having 62 observations. I think this is not of great concern. As a comparison, having only 41 observations Barth et al. (2006, p. 218) introduce the same amount of variables.

the intercept α and the slope β_1 on capital regulation while for the set of rich countries ($\text{Inc}_i = 1$) we estimate the intercept $\alpha + \zeta$ and the slope $\beta_1 + \delta$.

Private monitoring is more helpful under low corruption It turns out that in the estimation of equation 3 by OLS ζ and δ are both insignificant for all Basel pillars. I.e. no significant interaction is identified. The same is true when the income dummy is replaced by the corruption dummy. One exception is the interaction with private monitoring in 1999. In that case δ is significantly positive which means that for the sampled high corruption countries more private monitoring brings on average less stability than for low corruption countries. This pattern was blurred in the pooling over all countries in model 1 of table 7 where for 1999 the negative coefficient on private monitoring was not significant.[66]

As I found ζ of equation 3 to be insignificant one might drop it. Then one assumes that only the regression slopes differ between the groups. One might drop (only) δ as well, as it is insignificant, allowing just the intercepts to vary. Both approaches might lead to an omitted variable bias where variability is incorrectly assigned to the Basel pillars. Even so, models where only the intercept or the slopes are allowed to vary are explicitly described[67] such that I consider it an empirical question whether a random slope (fixed intercept) model is appropriate.

Firstly, I evaluate whether adding both the pure term and the interaction of Inc_i with a Basel pillar significantly increases the models' explanatory power. To do so I use an F-test[68] which considers in its test statistic the change in R^2 from adding regressors ("incremental F-test"),

$$F = \frac{(R_2^2 - R_1^2)/(k_2 - k_1)}{(1 - R_2^2)/(N - k_2 - 1)}$$

where index $i = 2$ stands for the model with the greater and index $i = 1$ for the model with the fewer number of predictors, R_i^2 is the R^2 of model i, k_i

[66] The significant interaction does not mean that private monitoring has a statistical effect for the subgroups. Estimating the equations for 1999 separately for $\text{Corruption}_i = 0$ and $\text{Corruption}_i = 1$ private monitoring is neither significant for low nor for high corruption countries.

[67] The approach can be described by a random coefficient regression model. Level 1 stands for the lowest level of aggregation which is the country level here; level 2 or the group level (Cohen et al., 2003, p. 545) is identified by the income or corruption dummy. If there is no variation in intercepts or slopes across the groups the random coefficient regression model is identical to (fixed) OLS regression (Cohen et al., 2003, p. 547).

[68] As far as the significance of changing one regressor is concerned a t-test would be sufficient. Then I still use the F test to limit the concepts introduced.

the number of predictors and N the number of observations in the model.[69] Applying the test I assume to find the most appropriate model by maximizing the share of total variation explained. In 1999, I find that adding both Inc_i and Corruption_i interacted with capital regulatory stringency significantly increases the model fit as the statistic $F = \frac{(0.56-0.48)/(10-8)}{(1-0.47)/(62-10-1)} = 5.88$ lies above the critical value $F_{10-8;62-10-1;1-0.01} = 5.05$ such that the Null of a population adjusted R^2 increment of zero is strongly rejected (p-value of 0.005). For 1999 and both Inc_i and Corruption_i – each introduced as pure term and as interaction with the Basel pillars – I find that all p-values for the incremental F-test lie below 0.012; for 2002 below 0.024; for 2006 above 0.097. For 2011 the p-values are not of interest as in all six variants the adjusted R^2 remains negative. In other words, introducing the interactions increases the model explanatory power considerably for 1999 and 2002 but not in later years.

Appropriateness of including only the interaction Still, when both pure and interaction term are included (equation 3) the coefficients are insignificant with the single exception noted above. Thus, secondly I consider dropping Inc_i and Corruption_i, respectively (ζ in equation 3) while leaving the interaction of the dummy with one of the Basel pillars in the model. This procedure is adequate as long as the model fit does not decrease significantly. Note, however, the alternative to drop the interaction (δ in equation 3) and just use Inc_i and Corruption_i which is the simpler model. I apply the decision rule to (a) use the simpler model unless the increase from including the interaction is at least 5% higher than the increase from including only Inc_i or Corruption_i; (b) additionally, the resulting model is only printed in table 8 and 9 when the interaction is significant and the incremental F-tests does not judge the full model (equation 3) as better ($p < .10$).

For survey III in 2006 no interaction is significant and for survey IV in 2011 all adjusted R^2 are negative. Table 9a gives the baseline model first. Model 2 shows an “interaction only surplus” of 18%. This is the increase in adjusted R^2 from adding $\text{CapitalRegulation} * \text{Inc}_i$ over and above the increase from adding just Inc_i. Following part (a) of the decision rule equation 3 is estimated with Inc_i dropped. Model 3 in table 9a does not show the interaction with private monitoring because the “interaction only surplus” of -2% is below 5% such that the simpler model is preferred. Part (b) of the decision rule leads in no case to the rejection of the model found in the first step. Therefore table 8 for

[69] Jaccard & Turrisi, 2003, p. 11f

Table 8: OLS regressions with interactions based on income: 1999/2002

(a) For 1999 / survey I

	(1)	(2)	(3)	(4)
Capital Regulation	-0.201^{***}	-0.093	-0.167^{***}	-0.151^{**}
Private Monitoring	-0.156	-0.110	-0.113	-0.105
Official Supervisory Power	0.018	0.038	0.020	0.058
Entry Requirements	0.055	0.039	0.048	0.048
Restrictions on Activities	0.223^{***}	0.137^{**}	0.154^{**}	0.159^{***}
Capital Regulation * Inc		-0.144^{***}		
Inc			-0.805^{***}	
Supervisory Power * Inc				-0.078^{***}
Observations	62	62	62	62
Interaction only surplus	-	18%	-2%	7%
Adjusted R^2	0.376	0.482	0.466	0.473
F Statistic	5.597^{***}	7.316^{***}	6.915^{***}	7.077^{***}

(b) For 2002 / survey II

	(1)	(2)	(3)
Capital Regulation	-0.047	-0.018	-0.017
Private Monitoring	-0.197^{**}	-0.118	-0.123
Official Supervisory Power	0.039	0.023	0.057
Entry Requirements	0.117	0.096	0.100
Restrictions on Activities	0.108	0.040	0.040
Inc		-0.849^{***}	
Supervisory Power * Inc			-0.076^{***}
Observations	77	77	77
Interaction only surplus	-	-13%, -13%	11%
Adjusted R^2	0.192	0.291	0.302
F Statistic	3.251^{***}	4.467^{***}	4.648^{***}

*p<0.1; **p<0.05; ***p<0.01
Dependent variable is $log(\mathrm{NPL}_i)$. Model 1 is supplied for comparison with table 7 and the analog for 2002. "Interaction only surplus" divides *the percentage increase in adjusted R^2 going from the baseline model 1 to a model where* Inc_i *and the relevant Basel pillar are interacted* by *the percentage increase from adding only* Inc_i. $\mathrm{Inc}_i = 1$ for high income countries is based on the World Bank classification. Legal origin variables and the constant are suppressed.

Inc_i and 9 for $\mathrm{Corruption}_i$ present no model where both the dummy and the interaction with a Basel pillar is included.[70]

For both 1999 and 2002 the results show with a $p < 0.05$ that high income countries face a lower bank fragility and corrupt countries face a higher bank

[70] Out of the models investigated the incremental F-test going from model 3 in table 9a to the full model is rejected with a p-value of 0.12 which is by far the weakest rejection under investigated models. Private monitoring interacted with $\mathrm{Corruption}_i$ was the only case where for the full model dummy or interaction were significant.

Table 9: OLS regressions with interactions based on corruption: 1999/2002

(a) For 1999 / survey I

	(1)	(2)	(3)	(4)
Capital Regulation	-0.201^{***}	-0.177^{**}	-0.171^{**}	-0.173^{**}
Private Monitoring	-0.156	-0.175	-0.208^{**}	-0.170
Official Supervisory Power	0.018	0.043	0.047	0.029
Entry Requirements	0.055	0.051	0.052	0.049
Restrictions on Activities	0.223^{***}	0.191^{***}	0.195^{***}	0.190^{***}
Corruption		0.760^{**}		
Private Monitoring * Corruption			0.103^{***}	
Supervisory Power * Corruption				0.074^{***}
Observations	62	51	51	51
Interaction only surplus	-	0%	9%	5%
Adjusted R^2	0.376	0.540	0.556	0.549
F Statistic	5.597^{***}	7.522^{***}	7.943^{***}	7.752^{***}

(b) For 2002 / survey II

	(1)	(2)	(3)
Capital Regulation	-0.047	-0.060	-0.061
Private Monitoring	-0.197^{**}	-0.237^{**}	-0.262^{***}
Official Supervisory Power	0.039	0.022	0.018
Entry Requirements	0.117	0.078	0.074
Restrictions on Activities	0.108	0.101	0.097
Corruption		0.766^{**}	
Private Monitoring * Corruption			0.097^{***}
Observations	77	61	61
Interaction only surplus	-	-4%, 0%	9%
Adjusted R^2	0.192	0.268	0.275
F Statistic	3.251^{***}	3.435^{***}	3.525^{***}

*p<0.1; **p<0.05; ***p<0.01
Compare notes in table 8. "Interaction only surplus" is calculated analog using Corruption_i. $\text{Corruption}_i = 1$ stands for corrupt countries.

fragility.[71] In 1999, more stringent capital regulation is stabilizing in high income countries while it has no effect in low income countries.[72] Both in 1999 and 2002 supervisory power is significantly more helpful in rich countries. The results for 1999 in table 9a show that capital stringency has a stabilizing effect regardless of the corruption level; this appears reasonable. In corrupt societies supervisory power is harmful from a theoretical perspective which is found in the data for 1999 but not in later years.[73] The interaction for corrupt countries

[71] Model 2 in table 8a/8b and model 2 in 9a/9b.
[72] Compare model 2 in 8a.
[73] Compare model 4 in 9a. Note that the effect of official supervision is nonnegative for low corruption countries as well.

and private monitoring has a positive coefficient, i.e. in corrupt countries private monitoring is to a smaller extent associated with higher stability. The significant coefficient on private monitoring of $-.21$ in 1999 (model 3 in 9a) shows that private monitoring decreases fragility in low corruption countries. The net effect for high corruption countries is numerically $-.21 + .10 = -.11$. A Wald test for the joint hypothesis that the coefficient on private monitoring and the coefficient on the interaction with private monitoring are both zero is rejected with $p < 0.01$. Thus, more private monitoring increases stability in corrupt countries as well, however to a smaller extent.[74] For 2002 the same result is obtained.

In conclusion, based on the cross-section (a) for the period around the financial crisis starting in 2007 we cannot identify a clear impact of regulation on bank stability (surveys III and IV in 2006 and 2011), (b) in 1999 and 2002 bank stability is higher in rich countries and lower in corrupt countries, (c) in 1999 and partly in 2002 capital regulation and supervisory power are more effective in rich countries to increase stability, (f) in 1999 and partly in 2002 private monitoring and supervisory power are less helpful in corrupt countries, (d) in 1999 and 2002 private monitoring significantly increases stability in corrupt countries, and (e) in 1999 supervisory power significantly decreases stability in corrupt countries.

4 Modeling the time dimension

A small amount of observations could be the reason that we hardly see a pattern of regulation on bank fragility in the cross-section for 2006 and 2011. This section introduces the time dimension to use as many observations as possible investigating the Basel pillars' effect on bank fragility. NPL are shown to exhibit a persistent pattern which reduces the confidence in the simple OLS cross-section results.[75] Statistically one can account for the dynamic panel structure by modeling the error term using an autoregressive process. Reproducing Boudriga et al. (2009) in subsection 4.1 follows this idea using panel corrected standard errors. Evidence is gained that serial correlation remains in the error terms casting doubt on the results in Boudriga et al. (2009). This is the motivation to explicitly model the dynamic process with lagged NPL as an

[74] When I estimate the model separately for $\text{Corruption}_i = 0$ and $\text{Corruption}_i = 1$, private monitoring is not significant. As noted above, separate estimation uses less information and thus leads to less significance.

[75] Bond, 2002, p. 141f

additional regressor in subsection 4.2. This section is based on a more comprehensive set of control variables which on the one hand limits the comparability with the cross-section analysis. On the other hand, we gain insights about the robustness of my findings from the different approach to model NPL.[76]

4.1 Estimation using panel-corrected standard errors (PCSE)

Boudriga et al. (2009) model NPL in a panel for the years 2002-06. Their approach is more involved than the work of Barth et al. because (a) they incorporate the time dimension, (b) they include a more realistic set of control variables, and (c) they try to avoid potential endogeneity by lagging time variant control variables. Up to now I assumed that NPL are only affected by the regulatory variables and the countries' legal origin. However, it seems to be obvious that the share of NPL is influenced by the current state of the economy. Nations' GDP growth is one of the control variables used in Boudriga et al. (2009). Their model can be written as

$$\begin{aligned} \mathrm{NPL}_{it} &= \alpha + \mathrm{regulation}'_{it}\,\beta + \mathrm{control}_i\,\gamma + \mathrm{control}'_{i,t-1}\,\delta + u_{it} \qquad (4) \\ & i = 1, \ldots, N,\ t \in (2002 - 2006),\ u_{it} \sim \mathrm{N}(0, \sigma_{\mathrm{i}}^2) \end{aligned}$$

where the regulatory variables vary over time and are thus written with a time index t. Based on the time frame 2002-06 surveys II and III of Barth et al. (2013b) are relevant. The NPL are available on a yearly frequency while survey II and III for 2002 and 2005/06, respectively. The question is how to merge the data. Boudriga et al. (2009, p. 294) attach the survey II values to the years 2002-04 and those for survey III to 2005-06. It is assumed that the variation in the regulatory setting between the surveys is minor. Definitions for the variables used in Boudriga et al. (2009) are shown in table 10. Variables already introduced in section 3.1 are omitted in the table. As control variables the authors use a time-invariant dummy capturing the countries' financial development (effect on NPL measured by coefficient γ) and four time-variant variables (effect on NPL measured by coefficients δ). The time-variant variables are capital to risk-weighted assets, bank provisions to NPL, return on assets (ROA) and the nominal GDP growth. The level of NPL might induce changes in these time-variant control variables. E.g. as a reaction to observing a higher share of NPL provisions to NPL might be increased. That would render right-hand side

[76] Wooldridge (2000, p. 625f) makes the point that a relationship should be significant in different models to be judged as robust.

variables dependent. To circumvent this endogeneity problem the time-variant control variables are lagged by one period.[77]

Boudriga et al. is not reproducible The time-invariant control variable is based on a financial development index and gives a 1 for developed countries and a 0 else. The variable is highly significant across the Boudriga et al. regressions in which they state that they have data for 59 countries[78,79] Notably, the source given, World Economic Forum (2008), contains a ranking only for 52 countries. The authors do not note that they improved the data, e.g. that they determined the ranking for more countries. I can neither explain nor cure this discrepancy making it impossible to reproduce the results in Boudriga et al. (2009).[80]

A second discrepancy in the data is minor compared to the missingness described. Boudriga et al. use a variable where required minimum capital is deducted from “Bank regulatory capital to risk-weighted assets (%)”. The source cited in Boudriga et al. (2009, p. 295) contains only “Bank Regulatory Capital to Risk-Weighted Assets”[81] and I was not able to find a source for the required minimal capital for the comprehensive country set and time frame used.[82]

Although the reproduction of Boudriga et al. (2009) is not possible, estimating a similar model allows insights about the robustness of the patterns found. Equation 4 pools the coefficients over countries and years. Boudriga et al. use panel-corrected standard errors (PCSE) in their estimation. In the regressions so far we controlled for heteroscedasticity relying on White (1980)’s standard errors. These are not appropriate in the panel setting[83] where the errors possibly exhibit panel structure. With panel structure Beck and Katz refer to (a) errors of different countries which are linked for the same point in time – contemporaneous correlation – e.g. when adjacent countries suffer under the same shock, (b) error variances which differ by country – cross-section heteroscedas-

[77] The implicit assumption is that changes in NPL are not informative two periods ahead. Lagging two periods or more, however, might prevent a recent trend to be recognized as driving the NPL around the same time. Using one lag appears to be a good compromise.

[78] Data on 59 countries for 2002-06 are $5 * 59 = 295$ observations.

[79] E.g. Boudriga et al., 2009, p. 299

[80] The countries for which the ranking is provided are different from those for which complete cases are available in Barth et al. (2013b). Using the World Economic Forum ranking variable in my re-estimation reduces the observations to 39, a major deviation from 59 in the original. In 2008, the ranking was produced for the first time. The data from the following survey is barely more comprehensive (55 instead of 52 countries are ranked; World Economic Forum, 2009).

[81] IMF, 2007, p. 168f

[82] However, I see the level of required minimum capital as rather alike at least over regions such that the difference might have no major impact.

[83] Beck & Katz, 1995, note 13

Table 10: Variable definitions and data sources for Boudriga et al. (2009)

Variable	Definition	Source
Bank regulatory capital to risk-weighted assets (%)	The capital adequacy of deposit takers. It is a ratio of total regulatory capital to its assets held, weighted according to risk of those assets. Note that due to differences in national accounting, taxation, and supervisory regimes, these data are not strictly comparable across countries.	IMF (2007) received by The World Bank (2013a)
Bank Provisions to nonperforming loans (%)	Loan loss provisions as a share of nonperforming loans. Nonperforming loans are loans for which the contractual payments are delinquent, usually defined as (...) being overdue for more than a certain number of days. Note that due to differences in national accounting, taxation, and supervisory regimes, these data are not strictly comparable across countries.	IMF (2007) received by The World Bank (2013a)
Bank Return on Assets	Data definitions follow, to the extent possible, the methodology of the Financial Soundness Indicators Compilation Guide. "This FSI is intended to measure deposit takers' efficiency in using their assets. It is calculated by dividing net income before extraordinary items and taxes by the average value of total assets (financial and nonfinancial) over the same period." (IMF, 2006, p. 184)	IMF (2008), IMF (2011), IMF (2013)
Foreign-Owned Banks	The extent to which the banking system's assets are foreign owned.	Barth et al. (2013b)
Bank Concentration (Asset)	The degree of concentration of assets in the 5 largest banks.	Barth et al. (2013b)
GDP growth (annual %)	Annual percentage growth rate of GDP at market prices based on constant local currency. Aggregates are based on constant 2005 U.S. dollars. GDP is the sum of gross value added by all resident producers in the economy plus any product taxes and minus any subsidies not included in the value of the products. It is calculated without making deductions for depreciation of fabricated assets or for depletion and degradation of natural resources.	The World Bank (2014b)
FDI rank dummy	Dummy variable taking 1 for countries with a financial development index numerically higher to the median rank and 0 otherwise. The highest developed country is ranked first such that the value 1 identifies relatively developed countries.	World Economic Forum (2008)

ticity[84] – e.g. when developing countries have more unexplained movement in NPL, and (c) error variances which differ for a country over time – panel-level heteroscedastic errors – e.g. due to unexplained changes in countries' debtors behavior. PCSE correct the standard errors for the panel structure and allow for correct inference. The key assumption needed for PCSE is the absence of autocorrelation in the individual error processes, i.e. no serial correlation.[85]

Serial correlation in PCSE The Wooldridge test for serial correlation based on first differences is used to check the absence of serial correlation.[86] Residuals are taken from an OLS regression based on equation 4 where all left- and right-hand variables are in first differences, no constant is specified and the covariance matrix is estimated robustly. The residuals are then regressed on their first lag estimated in the same robust way ($resid_{it} = \rho * resid_{i,t-1} + error_{it}$). Based on a Wald test for the simple linear hypothesis $H_0 : \rho = -.5$ a p-value < 0.01 is obtained which is strong evidence for serial correlation in the data.[87,88]

A strategy to get rid of the serial dependence is to include the lagged dependent variable as a regressor. This leads to endogeneity and motivates the application of linear dynamic panel estimators in subsection 4.2. Another option is to model the error terms with an AR1 autocorrelation structure. The Stata command *xtpcse* offers such an option.[89] Whether the errors are in fact free of serial correlation is not tested by Boudriga et al. To use the serial correlation test based on first differences we would need to incorporate the AR(1) error structure into the auxiliary regression. I prefer a test which rests upon the regression residuals $\hat{u}_{it}$ of the original regression. As PCSE is a pooled OLS approach with corrected standard errors I use the Wooldridge test for serial correlation after pooled OLS presented in Wooldridge (2010, p. 198f). An OLS regression of the form $NPL_{it} = \rho * \hat{u}_{i,t-1} + regressors + error_{it}$ is used where the lagged residual and the same regressors as in the original equation are included. Based on

[84] Beck & Katz, 1995, p. 636 call this "panel heteroscedasticity" which can be confused with panel-level heteroscedastic errors.

[85] Beck & Katz, 1995, p. 638 and note 13

[86] Wooldridge, 2010, p. 319f

[87] Intuitively $\rho = -.5$ under the Null because Δu_{it} and $\Delta u_{i,t-1}$ – where Δ is the first difference operator – share half of their elemens, the first containing $-u_{i,t-1}$ and the second $+u_{i,t-1}$.

[88] The Stata command xtserial (Drukker, 2003) can be reproduced using in both auxiliary OLS regressions the option cluster() which allows for heteroscedastic error terms of a country and is robust against serial correlation. These options allow for correct inference under structure in the errors when they deviate from the ideal assumptions. However, the errors are not transformed as then ρ would be a performance measure of the transformation.

[89] Controlling both for contemporaneous correlation across panels and heteroscedastic errors within panels – i.e. for a country over time – is not possible due to insufficient data. Thus, it is only controlled for the latter error term structure. Boudriga et al. (2009) do not explicitly state that they correct for serial correlation but possibly they set the option *correlation(ar1)* of the Stata command *xtpcse* as well.

robust standard errors the hypothesis $H_0 : \rho = 0$ stands for no serial correlation and is rejected with a p-value of $p < 0.01$.

As I find a strong violation of the absence of serial correlation assumption when estimating the model in equation 4 by PCSE I consider it as inappropriate to interpret these results based on standard errors while the coefficients themselves should be consistent. Having said that, I provide the re-estimation of Boudriga et al. (2009, p. 299) in the appendix in table 13.[90]

4.2 Estimation using a dynamic panel

Serial dependence is a core challenge in explaining bank fragility in terms of NPL. The strategy to adjust the error structure using PCSE does not appear to be successful, at least for the data I use. An alternative to cope with serial dependence is to include the lagged NPL as a regressor. This allows to model a dynamic process where current realizations are influenced by past ones. I allow for individual specific effects μ_i to capture countries' heterogeneity. Writing the model

$$\begin{aligned} log(\mathrm{NPL}_{it}) \quad &= \quad \alpha * log(\mathrm{NPL}_{i,t-1}) + \mathrm{regulation}_i'\,\beta \\ &\quad + \mathrm{control}_i'\,\gamma + \mathrm{control}_{i,t-1}'\,\delta + \mu_i + v_{it} \\ &\quad i = 1, \dots, N,\ for\ t \in (2002-2006), t \in (2007-2011), \\ &\quad E[\mu_i] = E[v_{it}] = E[\mu_i v_{it}] = 0,\ endogenous\, are\ \mathrm{NPL} \end{aligned} \tag{5}$$

for the period $t-1$, i.e. $log(\mathrm{NPL}_{i,t-1}) = ... + \mu_i + v_{i,t-1}$, shows that the lagged dependent and the country specific effects μ_i are correlated and thus an endogeneity bias is present.[91] Using first differences to eliminate the time-invariant μ_i is a natural approach. However, $log(\mathrm{NPL}_{it}) - log(\mathrm{NPL}_{i,t-1}) = \alpha * (log(\mathrm{NPL}_{i,t-1}) - log(\mathrm{NPL}_{i,t-2})) + \delta * ... + v_{it} - v_{i,t-1}$ contains on the right-

[90] Most of the significant variables are numerically near and the levels of significance basically match. Capital to risk-weighted assets is the variable which differs conceptually and is insignificant while relevant at the 10%-level in the original. In my regression ROA is significant while it is not in the original. Dropping the problematic FDI rank variable increases observations to 192-200 and takes all significance from foreign-owned banks. The rest of the results are only marginally influenced; supervisory power in model 3 turns significant at the 1% level.

[91] For a large time dimension T the endogeneity can be irrelevant. The data here is of small T, large N. If a countries' financial crisis is not modeled and thus captured by the error term all else equal the higher NPL_{it} will be captured by a greater fixed effect. For $t+1$ both the fixed effect and the regressor $NPL_{i,t-1}$ are larger and thus endogeneity is present. The effect is relevant when the shock affects μ_i strongly which is the case for the small T here (Roodman, 2009, p. 101).

hand side the negative $v_{i,t-1}$ and a positive $v_{i,t-1}$ in the lagged term.[92] Due to the negative correlation in the error term both OLS on first differences and a fixed effects estimator lead to an underestimation of α and to misleading results. Using pooled OLS on the undifferenced equation 5 leads to an overestimation of α because the positive μ_i is contained in $log(\mathrm{NPL}_{i,t})$ and in $log(\mathrm{NPL}_{i,t-1})$. Bond (2002, p. 144) proposes to use these bounds as a check on results of GMM estimators I introduce below.

I follow the variable selection in Boudriga et al. (2009) where I replace the financial development dummy – which is not available in an adequate scope – by the income dummy introduced in section 3.5.[93] I continue to model $log(\mathrm{NPL}_{it})$ as dependent because it produced more normal errors (compare figure 3). I put in the three Basel pillars at once because the pillars might interact and the panel setting allows the identification of the higher amount of coefficients.[94] For 2002-06 I obtain 0.52 and 0.98 as bounds for $\hat{\alpha}$.

To handle the endogeneity in equation 5 I use instrumental variables (IVs) implemented by the generalized method of moments (GMM). GMM models of the 1980s relied on the assumption that regressors are strictly exogenous, or alternatively, strictly exogenous IVs are available. Arellano and Bond (1991) propose a GMM estimator with the more restrictive assumption of absence of serial correlation in the error terms which allows to cope with regressors and IVs that are not strictly exogenous. In their difference GMM estimator unobservable individual specific effects μ_i are differenced out. Then lags of the time-variant regressors are used as instruments to handle endogeneity. When there is no serial correlation in the errors of the differenced equation 5 the endogenous regressor $log(\mathrm{NPL}_{i,t-1}) - log(\mathrm{NPL}_{i,t-2})$ can be instrumented with its second and deeper lags.[95,96]

Weights used to calculate the variance of the GMM estimators can be based on a matrix that does not depend on estimated parameters (one-step GMM) or on an initial consistent estimate which enters the weight matrix (two-step GMM).

[92] The differenced lagged term is the equation written for the period $t-1$, i.e. $log(\mathrm{NPL}_{i,t-1}) - log(\mathrm{NPL}_{i,t-2}) = \alpha * (log(\mathrm{NPL}_{i,t-2}) - log(\mathrm{NPL}_{i,t-3})) + \delta * ... + v_{i,t-1} - v_{i,t-2}$.

[93] The FDI rank dummy in Boudriga et al., 2009 (compare table 10) is highly significant over the authors' regressions but is only available for a little amount of countries. The high income dummy based on the World Bank classification is clearly different conceptually but to a less extent concerning the effective country grouping and captures an important country characteristic as well.

[94] A possible influence of the Basel pillars on each other is assumed to be irrelevant over all the regressions in Boudriga et al. (2009).

[95] Arellano & Bond, 1991, p. 277f

[96] The importance to check the assumption that $\rho = 0$ in $\hat{v}_{it} = \hat{\rho}\hat{v}_{i,t-2} + error_{it}$ is expressed by the title of the paper introducing difference GMM, namely "Some tests of specification for panel data".

The uncertainty from the initial estimate translates into a bias of the efficient two-step GMM in small samples which can be accounted for by the Windmeijer (2005) correction.[97] However, the two-step GMM is only better under certain assumptions for which it is unclear how to check them in practice.[98] The bias of the efficient two-step GMM is affected by the general concern of weak instruments as well as by the use of too many instruments.[99] The number of instruments has to be adjusted to the sample size[100] while it is unclear what amount of instruments is too high.[101] I limit the lags used as IVs for the endogenous variable as much as possible given the rather small amount of observations.

Model 1 in table 11 gives the difference GMM estimator for equation 5 where the lags used as instruments for the endogenous variable NPL_{it} are limited to two.[102] The choices in terms of one- and two-step and the use of t instead of z statistics due to the small sample do not qualitatively affect the results.[103] The estimated coefficient on the lagged dependent ($\hat{\alpha}$) is outside the credible range and extraordinarily high.

Blundell and Bond (1998, p. 120ff) show that under small T and a "moderately large"[104] autoregressive parameter α difference GMM exhibits low precision and a considerable bias in small samples. In other words, the approach to instrument first differences of regressors with the levels of the lagged regressors is not helpful here.[105] The idea of system GMM is to introduce a second equation where the regressors are in levels and sufficiently lagged first differences of the regressors

[97] Windmeijer, 2005, p. 27ff

[98] A symmetric finite sample distribution of the GMM estimator (Windmeijer, 2005, p. 31) and moment conditions which are linear in the parameters (Windmeijer, 2005, p. 29) are assumed.

[99] Windmeijer, 2005, p. 31

[100] Arellano & Bover, 1995, p. 41

[101] Roodman, 2009, p. 99

[102] I follow the advice in Roodman (2009, p. 129f) about GMM specification choices to report. All variables not noted to be endogenous – in equation 5 only NPL – are assumed to be exogenous, or "IV-style instruments". In all GMM estimations as additional IV-style instruments control of corruption – "perceptions of the extent to which public power is exercised for private gain" – and government effectiveness – "perceptions of (...) the quality of the civil service and the degree of its independence from political pressures" – are added (The World Bank, 2014c).

[103] I always use the robust option of the *xtabond2* command in Stata which yields standard errors which are consistent against any pattern of heteroscedasticity and autocorrelation within the countries in the one-step procedure. The two-step procedure is robust by construction while the robust option triggers the Windmeijer correction.

[104] Blundell & Bond, 1998, p. 115

[105] Blundell and Bond (1998, p. 134) present evidence for the clear superiority of system GMM in the case of $\alpha = .8$, $N = 200$, and $T = 11$. In model 1 of table 11 all three parameters are worse indicating the use of system GMM.

Table 11: Dynamic panel estimation for 2002-2006

	(1)	(2)	(3)	(4)
Log of NPL (L)	1.16***	0.95***	0.97***	0.92***
Capital Regulation	0.03	0.00	0.00	−0.01
Private Monitoring	−0.15**	−0.03	−0.04*	−0.05*
Supervisory Power	0.07*	0.02	0.01	0.01
Capital to RWA (L)	0.01	−0.01	−0.00	−0.01
Prov. to NPL (L)	0.01**	0.00	0.00	0.00
ROA (L)	−0.04	−0.01	0.00	−0.03
Government-Own.	−0.88	−0.21*	−0.29**	−0.22
Foreign-Own. Banks	−0.76*	−0.02	−0.05	−0.01
Asset concentr.	0.05	−0.11	−0.18	−0.28
GDP growth (L)	0.00	0.00	−0.00	0.00
Inc		−0.24	−0.22	−0.26
Observations	122	190	190	190
Countries	58	68	68	68
GMM method	difference	system	system	system
Steps	two-step	one-step	two-step	two-step
Instruments	18	23	23	33
Arellano-Bond test	0.53	0.88	0.86	0.56
Sargan test (p-value)	0.36	0.45	0.45	0.81
Hansen test (p-value)	0.35	0.45	0.45	0.54
Assumed endogenous	NPL	NPL	NPL	NPL, ROA

* p<.1, ** p<.05, *** p<.01
Dependent variable is $log(\text{NPL}_{it})$. The constant is suppressed. The Arellano-Bond test has the H_0 of no autoregressive process of 2^{nd} order in first differences. The Sargan and the Hansen test have the Null of strictly exogenous instruments.

are used as instruments.[106] The identification of the level equation relies on the availability of variables that have a constant correlation with the country specific effects μ_i.[107] Given this additional assumption of system GMM the estimator is shown to considerably increase precision and reduce finite sample bias.[108]

Model 2 in table 11 gives the one-step system GMM estimator for equation 5. The coefficient of the lagged dependent is now in the credible range, 23 instruments on 68 countries appear reasonable, the Arellano and Bond (1991) serial correlation test on the first difference equation does not reject the Null of absence of an autoregressive process of order 2 (p-value of 0.88), and the instruments appear to be sufficiently exogenous.[109] Model 3 is based on the two-step procedure for which the assumptions are a concern (see above). The

[106] Arellano & Bover, 1995, p. 45
[107] Arellano & Bover, 1995, p. 44
[108] Blundell & Bond, 1998, p. 133f
[109] The p-value of the Sargan test with the Null of exogenous instruments is 0.45.

specification tests are passed similar well as in model 2 while there is more significance. For the cross-section and 2002 I found evidence that both private monitoring and higher income are associated with higher stability (table 8b and 9b). Under a different set of control variables for 2002-06 we again find private monitoring to increase bank stability (p-value of 0.08). Controlling for the income[110] – which appeared crucial in the cross-section – model 3 gives us the idea that a higher share of government-owned banks reduces bank fragility.

Model 4 in table 11 shows the impact of introducing year dummies and assuming $\text{ROA}_{i,t-1}$ to be endogenous additionally to NPL. Year dummies model a potential pattern over time.[111] I suspect that a higher share of NPL carries information about the lagged ROA and the lagged provisions to NPL. Both variables are lagged in Boudriga et al. (2009) to prevent endogeneity. Their underlying idea is that NPL today influences both variables; e.g. the ROA decrease when more loans are in default. The dependent variable then affects the right-hand side. My point is that one lag might not be sufficient to prevent endogeneity under the observed autocorrelation in NPL_{it}. Rather, $log(\text{NPL}_{it})$ correlates with $log(\text{NPL}_{i,t-1})$ and thus NPL today carry information about $\text{ROA}_{i,t-1}$. The validity of the instruments in model 4 increases from modeling ROA as endogenous. Private monitoring is found to robustly decrease fragility before the financial crisis while the share of government banks is not a robust driver of NPL.[112]

For the financial crisis period 2007-2011 table 12 gives the results.[113] As the credible range for $\hat{\alpha}$ I obtain 0.60-0.91.[114] Model 1 is analog to the one in table 11 shown for the one-step procedure and qualitatively unchanged using two-step. The autoregressive coefficient $\hat{\alpha}$ is outside the credible range, here it is too small. The serial correlation test for AR(2) gives a $p = 0.15$ which is no strong evidence for the absence of serial correlation in the errors. Hence, time

[110] Note that difference GMM is not able to identify time-invariant regressors.

[111] In equation 5 we introduce a term λ_t representing time fixed effects. Dummies for $T-1$ periods are included. Applying the difference GMM estimator, Arellano & Bond, 1991, p. 288 include time fixed effects as well.

[112] The residuals of model 3 and 4 in table 11 appear rather randomly distributed. Higher NPL might induce an increase in supervisory power as well. Modeling supervisory power as endogenous drives up the instrument count to 37 and the coefficient on private monitoring turns insignificant. The stabilizing effect of private monitoring is thus not strictly robust but depends on the assumptions one is willing to make.

[113] I have to decide how to attach survey III and IV to the years 2007-11. For 2002-06 I followed Boudriga et al. (2009, p. 294). I attach survey III representing 2005-06 to 2005-08 while survey IV representing 2011-12 to 2009-11 assuming that regulation is partly anticipated.

[114] The credible range for 2007-11 is calculated including year dummies. For 2002-06 I excluded year dummies in the auxiliary regressions as they were not needed to prevent serial correlation. Excluding year dummies in 2007-11 yields a marginally different credible range of 0.58-0.89.

fixed effects are introduced. This yields $\hat{\alpha} = 0.85$ and 0.95 for the one-step and two-step procedure, respectively. The autoregressive parameter is high. This motivates the use of system GMM in model 2 where the regression results are invalid because both the Sargan and the Hansen (1982, p. 1049f) test reject the validity of the instruments. The Sargan test becomes inconsistent when the errors deviate from the ideal. The Hansen test is more consistent under deviations in the one-step estimation but is weakened by a high number of instruments.[115]

Table 12: Dynamic panel estimation for 2007-2011

	(1)	(2)	(3)
Log of NPL (L)	0.19	0.92***	0.91***
Capital Regulation	0.06	0.02	0.02
Supervisory Power	0.00	0.00	0.01
Private Monitoring	0.05	0.04**	0.03
Capital to RWA (L)	0.00	−0.00	0.00
Prov. to NPL (L)	−0.00**	0.00	0.00
ROA (L)	−0.02	−0.02	−0.13***
Government-Own.	2.16	−0.19	−0.24
Foreign-Own. Banks	0.20	0.08	0.07
Asset concentr.	−0.40	0.08	0.09
GDP growth (L)	0.01	0.00	0.01*
Inc		−0.01	−0.04
Observations	251	283	283
Countries	76	77	77
GMM method	difference	system	system
Steps	one-step	one-step	one-step
Instruments	22	32	46
Arellano-Bond test	0.15	0.37	0.38
Sargan test (p-value)	0.004	0.000	0.000
Hansen test (p-value)	0.006	0.002	0.066
Assumed endogenous	NPL	NPL	NPL, ROA

* p<.1, ** p<.05, *** p<.01
Dependent variable is $log(\text{NPL}_{it})$. Year dummies and constant are suppressed. Model 1 and 2 are invalid as their Sargan and Hansen tests reject the instruments' validity.

One might check the effect of all kinds of additional instruments on validity. I restrict myself to test for the already included ROA and the provisions to NPL whether they are actually exogenous. Model 3 in table 12 is the same as model 2 where the $\text{ROA}_{i,t-1}$ are now assumed to be endogenous. One might trust these results because the autoregressive coefficient is (just) in the credible range, serial correlation does not seem to be an issue ($p = .38$) and the p-value $p = .07$ of the Hansen test does not reject the Null of valid instruments at a 5% level

[115] Hansen, 1982, p. 1049 and Roodman, 2009, p. 97f

of uncertainty. Critical is model 3 because the additional endogenous regressor drives up the instrument count to 46 having observations for 77 countries which is an instrument share that raises a concern about the validity of the Hansen test. Following the Sargan test the validity of the instruments is rejected and the results should not be trusted.

Being aware of the limitations of model 3 in table 12 the crisis period is characterized by bank fragility not influenced by any of the Basel pillars. Instead, another pattern emerges where higher ROA reduce bank fragility on average over all countries. A higher GDP growth should reduce the NPL as well as it becomes easier to pay credit rates under higher income and nominally fixed rates. However, GDP growth is with a numerically rather small coefficient linked to higher fragility. This could represent NPL soaring with the financial crisis especially in the countries where a credit boom increased the GDP more, i.e. a positively linked effect. The explanation assumes that NPL affects the lagged GDP growth in the regression which is a sign that the first lag of GDP is not enough to prevent endogeneity. As we are clearly limited in the number of observation it does not appear reasonable to model GDP growth as another endogenous variable. Rather, I have to assume that the potential endogeneity of the variable does not affect the coefficients of interest strongly.

When I replace the dummy Inc_i in tables 11 and 12 by $\mathrm{Corruption}_i$ the coefficient on the dummy term remains insignificant. An interpretation is that the control variables added in equation 5 capture the important aspects of rich and corrupt countries such that the subsamples generated by $\mathrm{Corruption}_i$ and Inc_i no more differ in their level. Introducing the pure dummies and the dummies interacted with the Basel pillars yields no significant results for 2002-06. In the cross-section I was not able to obtain results for 2006 as well.[116] Based on the 2007-11 model in table 12 I do not investigate the effect of interactions as the model is at the threshold of being invalid.

In conclusion, based on the GMM estimation the years before the financial crisis appear as coined by the stabilizing effects of private monitoring. With the financial crisis starting in 2007 the Basel pillars can no more explain stability which is consistent with the cross-section results. Given doubt about the validity of the GMM approach in 2007-11 NPL are driven by ROA. Interactions were not found to be significant for 2002-06.

[116] For 2002 we were only able to identify differences in the Basel pillars impact on NPL when we applied the decision rule to use the interaction without the level dummy in some cases.

5 Conclusion

I empirically investigate how countries should implement the Basel accords to reform for the better. I focus on the effect of the Basel pillars on the stability of the banking sector. The World Bank data underlying the analysis requires aggregation. The indexation in Barth et al. (2013b) is shown to have considerable weaknesses and I clarify assumptions needed for an investigation based on the data. The extent to which bank fragility can be explained by the Basel pillars deteriorates over the time frame 1999-2011. Tackling endogeneity and the persistent pattern of nonperforming loans (NPL) using dynamic panel estimators shows that for 2002-06 private monitoring is associated with higher bank stability over the hole country set. Investigating country differences demonstrates that richer and less corrupt countries enjoy more stable banking systems. For 1999 and partly for 2002 I show that in corrupt countries stability is on average decreased when supervisory power goes up, and rises less than for incorrupt countries in increased private monitoring. For the financial crisis period 2007-11 the dynamic panel approach indicates that stability is higher with more productive banks while the Basel pillars do not have an impact.

Policy makers should take away from the study that for countries suffering from high corruption and low income (1) increasing supervisory power has a potentially destabilizing effect and (2) more private monitoring and capital regulation is less helpful than for the typical country engaged in developing the Basel accords.

These recommendations are cautiously phrased considering the limitations of the analysis. A multiplicity of distinct measures is subsumed under the headline of each Basel pillar. Even so supervisory power has downsides, concrete measures under supervisory review can be helpful in poor and corrupt countries. A direction for further research is a tailored compilation of indices using the unaggregated answers in the World Bank surveys. Thus packages of measures with stabilizing effects for developed and for corrupt countries can be identified better. The variable NPL limits the number of countries available in my analysis. The missingness in NPL could be non-random as the countries which are able to provide the data might have a more sophisticated banking system. This motivates the use of sample selection models where in a first step the probability of countries to participate is determined; or to choose a different proxy for bank fragility over which Gadanecz and Jayaram (2008) give an overview. In survey IV representing 2011/12 new questions were included as a reaction to the financial crisis. For comparability over time these questions are excluded in my analysis but are important for further research. Not only is "the systematic collection of data on bank regulatory and supervisory policies only in its nascent stages"[117]. Accordingly limited is the research on how to reform bank regulation for the better.

[117] Barth et al., 2013a, p. 1f

Appendix

Table 13: PCSE approach re-estimating Boudriga et al. (2009)

	(1)	(2)	(3)	(4)
Capital to RWA (L)	−0.01	−0.01	−0.05	0.00
Prov. to NPL (L)	−0.04***	−0.04***	−0.04***	−0.04***
ROA (L)	−1.13**	−1.13**	−1.38***	−1.24**
Government-Own.	1.88	1.89	1.47	1.61
Foreign-Own. Banks	−6.52**	−6.53**	−6.97**	−6.92**
Asset concentr.	−5.26**	−5.25**	−4.60**	−4.80*
GDP growth (L)	0.04	0.04	0.03	0.07
FDI rank	−6.44***	−6.44***	−6.33***	−6.87***
Capital Regulation		0.02		
Supervisory power			0.43**	
Private Monitoring				0.45
Observations	115	115	115	111
Adj. R^2	0.48	0.48	0.50	0.49

Note: *p<0.1; **p<0.05; ***p<0.01

Dependent variable is NPL which are modeled for 2002-06. PCSE are used (Beck & Katz, 1995) where the errors are modeled by an AR(1)-process. The validity of the approach is contested because the absence of serial correlation is rejected (compare section 4.1, page 34). The constant is suppressed.

References

Arellano, M., & Bond, S. (1991). Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations. *The review of economic studies*, *58*(2), 277–297.

Arellano, M., & Bover, O. (1995). Another look at the instrumental variable estimation of error-components models. *Journal of Econometrics*, *68*(1), 29–51.

Barth, J., Caprio, G., & Levine, R. (2006). *Rethinking Bank Regulation: Till Angels Govern.* Cambridge University Press.

Barth, J., Caprio, G., & Levine, R. (2012). The Evolution and Impact of Bank Regulations. *World Bank Policy Research Working Paper*(6288).

Barth, J., Caprio, G., & Levine, R. (2013a). Bank regulation and supervision in 180 countries from 1999 to 2011. *Journal of Financial Economic Policy*, *5*(2), 111-219. Available from http://faculty.haas.berkeley.edu/ross_levine/Papers/Bank_Regulation_and_Supervision_Around_the_World_15JAN2013.pdf (Page references relate to the version provided by the author Ross Levine under the link.)

Barth, J., Caprio, G., & Levine, R. (2013b, January). *Bank regulation and supervision in 180 countries from 1999 to 2011: Database.* Available from http://faculty.haas.berkeley.edu/ross_levine/Regulation.htm

Basel Committee on Banking Supervision. (2004, June). *International Convergence of Capital Measurement and Capital Standards: A Revised Framework.* Available from http://www.bis.org/publ/bcbs107.pdf

Basel Committee on Banking Supervision. (2010, December). *Basel III: A global regulatory framework for more resilient banks and banking systems.* Available from http://www.bis.org/publ/bcbs189.pdf

Beck, N. L., & Katz, J. N. (1995). What to do (and not to do) with Time-Series Cross-Section Data. *American political science review*, *89*(3), 634–647.

Blundell, R., & Bond, S. (1998). Initial conditions and moment restrictions in dynamic panel data models. *Journal of Econometrics*, *87*(1), 115–143.

Bond, S. R. (2002). Dynamic panel data models: a guide to micro data methods and practice. *Portuguese Economic Journal*, *1*(2), 141–162.

Bortz, J., & Schuster, C. (2010). *Statistik für Human-und Sozialwissenschaftler* (7th ed.). Springer.

Boudriga, A., Taktak, N. B., & Jellouli, S. (2009). Banking supervision and nonperforming loans: a cross-country analysis. *Journal of Financial Economic Policy*, *1*(4), 286–318.

Bring, J. (1994). How to Standardize Regression Coefficients. *The American Statistician*, *48*(3), 209–213.

Bühner, M. (2004). *Einführung in die Test- und Fragebogenkonstruktion.* Addison-Wesley Verlag.

Carmines, E. G., & Zeller, R. A. (1979). Reliability and validity assessment. In (Vol. 07-017). Sage.

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences* (3rd ed.). Lawrence Erlbaum Associates.

Cook, R. D. (1977). Detection of Influential Observation in Linear Regression. *Technometrics*, *19*(1).

Cook, R. D., & Weisberg, S. (1982). *Residuals and Influence in Regression.* Chapman and Hall.

Cortina, J. M. (1993). What Is Coefficient Alpha? An Examination of Theory and Applications. *Journal of Applied Psychology*, *78*(1), 98–104.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, *16*(3), 297–334.

DICE Database. (2014). *Corruption Perceptions Index, 1995-2013.* Available from www.cesifo-group.de/DICE/fb/pXBucnPs

Drukker, D. M. (2003). Testing for serial correlation in linear panel-data models. *Stata Journal*, *3*(2), 168–177.

Gadanecz, B., & Jayaram, K. (2008, July). Measures of financial stability – a review. In *Proceedings of the IFC Conference on "Measuring financial innovation and its impact", Basel* (Vol. 31, pp. 365–380).

Greene, W. H. (2008). *Econometric analysis* (6th ed.). Pearson Prentice Hall.

Hansen, L. P. (1982). Large sample properties of generalized method of moments estimators. *Econometrica*, *50*(4), 1029–1054.

IMF. (2006, March). *Appendix II. Summary of Guidance for Each Financial Soundness Indicator.* Available from http://www.imf.org/external/pubs/ft/fsi/guide/2006/pdf/appendix.pdf

IMF. (2007, October). *Statistical Appendix on Global Financial Stability Report: Financial Soundness Indicators.* Available from http://www.imf.org/External/Pubs/FT/GFSR/2007/02/pdf/statappx.pdf

IMF. (2008, April). *Global Financial Stability Report: Financial Soundness Indicators.* Available from http://www.imf.org/External/Pubs/FT/GFSR/2008/01/index.htm

IMF. (2011, April). *Financial Soundness Indicators: FSI tables.* Available from http://fsi.imf.org/fsitables.aspx

IMF. (2013, October). *Financial Soundness Indicators: FSI tables.* Available from http://fsi.imf.org/fsitables.aspx

Jaccard, J., & Turrisi, R. (2003). *Interaction Effects in Multiple Regression* (2nd ed.). Sage.

La Porta, R., Silanes, F. Lopez-de, & Shleifer, A. (2008). *The Economic Consequences of Legal Origins: Database* (Vol. 46) (No. 2). Available from http://scholar.harvard.edu/shleifer/publications

Long, J. S., & Ervin, L. H. (2000). Using Heteroscedasticity Consistent Standard Errors in the Linear Regression Model. *The American Statistician*, *54*(3), 217–224.

Mela, C. F., & Kopalle, P. K. (2002). The impact of collinearity on regression analysis: the asymmetric effect of negative and positive correlations. *Applied Economics*, *34*, 667–677.

Roodman, D. (2009). How to do xtabond2: An introduction to difference and system GMM in Stata. *Stata Journal*, *9*(1), 86–136.

Shapiro, S. S., & Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika*, *52*, 591-611.

Stigler, G. J. (1971). The Theory of Economic Regulation. *The Bell Journal of Economics and Management Science*, 3–21.

The World Bank. (2013a, November). *Global Financial Development Database (GFDD).*

The World Bank. (2013b). *Global Financial Development Report 2013: Bank Regulation and Supervision Survey.* Available from http://go.worldbank.org/WFIEF81AP0

The World Bank. (2014a). *How we Classify Countries.* Available from http://data.worldbank.org/about/country-classifications

The World Bank. (2014b). *World Development Indicators: database.* Available from http://data.worldbank.org/data-catalog/world-development-indicators

The World Bank. (2014c). *Worldwide Governance Indicators: database.* Available from http://data.worldbank.org/data-catalog/worldwide-governance-indicators

Transparency International. (2013). *Corruption Perceptions Index.* Available from www.transparency.org/policy_research/surveys_indices/cpi

Von Auer, L. (2007). *Ökonometrie: Eine Einführung* (4th ed.). Springer.

Wang, B. (2014). *bstats: Ordinary Regression Analysis with R.* Available from http://CRAN.R-project.org/package=bstats (R package version 1.1-11-4)

White, H. (1980). A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity. *Econometrica*, *48*, 817–838.

Windmeijer, F. (2005). A finite sample correction for the variance of linear efficient two-step GMM estimators. *Journal of Econometrics*, *126*(1), 25–51.

Wooldridge, J. M. (2000). Introductory Econometrics: A Modern Approach. *South-Western College Publishing, Australia*.

Wooldridge, J. M. (2010). *Econometric Analysis of Cross Section and Panel Data.* MIT press.

World Economic Forum. (2008). *Global Financial Development Report.* Available from http://www.weforum.org/pdf/FinancialDevelopmentReport/Rankings.pdf

World Economic Forum. (2009). *Global Financial Development Report.* Available from http://www3.weforum.org/docs/WEF_FinancialDevelopmentReport_2009.pdf